2023

Potential Alzheimer's Disease Plasma Biomarkers

Taylor Estepp
University of Kentucky, tges222@uky.edu
Digital Object Identifier: https://doi.org/10.13023/etd.2023.325

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Taylor Estepp, Student

Dr. Richard Charnigo, Major Professor

Dr. Erin Abner, Director of Graduate Studies

POTENTIAL ALZHEIMER'S DISEASE PLASMA BIOMARKERS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Public Health
at the University of Kentucky

By

Taylor Estepp

Lexington, Kentucky

Director: Dr. Richard Charnigo, Professor of Biostatistics

Lexington, Kentucky

2023

ABSTRACT OF DISSERTATION

POTENTIAL ALZHEIMER'S DISEASE PLASMA BIOMARKERS

In this series of studies, we examined the potential of a variety of blood-based plasma biomarkers for the identification of Alzheimer's disease (AD) progression and cognitive decline. With the end goal of studying these biomarkers via mixture modeling, we began with a literature review of the methodology. An examination of the biomarkers with demographics and other health factors found evidence of minimal risk of confounding along the causal pathway from biomarkers to cognitive performance. Further study examined the usefulness of linear combinations of biomarkers, achieved via partial least squares (PLS) analysis, as predictors of various cognitive assessment scores and clinical cognitive diagnosis. The identified biomarker linear combinations were not effective at predicting cognitive outcomes. The final study of our biomarkers utilized mixture modeling through the extension of group-based trajectory modeling (GBTM). We modeled five biomarkers, covering a range of functions within the body, to identify distinct trajectories over time. Final models showed statistically significant differences in baseline risk factors and cognitive assessments between developmental trajectories of the biomarker outcomes. This course of study has added valuable information to the field of plasma biomarker research in relation to Alzheimer's disease and cognitive decline.

KEYWORDS: AD, cognition, plasma biomarkers, PLS, GBTM

Taylor Estepp

06/25/2023

DEDICATION

To Scott, Tracy, and Sam.

ACKNOWLEDGMENTS

I could not have completed this dissertation without the help of many very important people. I am forever grateful for the support I received from those around me throughout this process. To my parents, thank you for instilling in me a good work ethic, for setting me up for success from the very beginning, and for always asking if I should be studying. To Sam, thank you for reminding me that I can do this and for encouraging me every day. To Hailee, thank you for getting me through the hard times and for saying yes to ordering pizza. To Cameron, Shaowli, and Michael, thank you for your support, your help, and mostly your friendship through the entirety of our grad school experience. To the rest of my friends and family, your love and support mean the world.

To my entire committee, thank you for the time and energy you sacrificed to help me make this happen. And a special thank you to Dr. Abner and Dr. Charnigo for your guidance, your flexibility, and above all your support. Your contributions helped better not only my work, but me as a scientist as well.

Chapter 2 included below is the original manuscript submitted to the Alzheimer's and Dementia Journal. The manuscript was subsequently accepted following revisions and additional data analysis, neither of which is included here. Thank you to each of my coauthors for helping me publish the best work we could. Wiley states "If you are the author of a published Wiley article, you have the right to reuse the full text of your published article as part of your thesis or dissertation. In this situation, you do not need to request permission from Wiley for this use." Citation for the published article follows: Estepp, TG, Charnigo, RJ, Abner, EL, et al. Associations of potential ADRD plasma biomarkers in cognitively normal volunteers. *Alzheimer's Dement*. 2023; 1-9. https://doi.org/10.1002/alz.13000.

Finally, thank you to Courtney Walker and Ximena Gonzalez for creating the template for this document.

TABLE OF CONTENTS

CHAPTER 1

Nonparametric Regression Review

Introduction

Before we dive into this review of nonparametric regression methods, it is important to understand what we mean by nonparametric. The word can take meaning in two very distinct ways. The perhaps more abstract definition of nonparametric describes a set of methods used to model the data when it cannot be described by a finite number of parameters. More commonly nonparametric statistics are defined to be statistical methods in which the data at hand is not required to fit into one of the parametric families of distributions [1]. Both definitions of the word are intuitive, and it is useful to note that they are not mutually exclusive.

There are many methods of nonparametric regression, along with many justifications for its use over (or even hand in hand) with the much more popular forms of parametric regression. In nonparametric regression we take our information from the data at hand as opposed to labeling it with a parametrized function. Immediately we can see that nonparametric regression will be of significant use if the data we have does not readily fit into a known family of parametric functions. Nonparametric regression methods arguably require more justification of use, as using less common and less well-known methods than the popular parametric methods and comes with the added burden of correctly conveying results and motivating readers to think outside their comfort zones. The added burden may not always be worth the trouble [2], but nonparametric regression is vastly underused in situations where it would be greatly helpful. The hope of many is that one day nonparametric statistics will be as common as its parametric counterpart [3].

Motivation

Section 1.1 of Hardle (1990) [4] presents many of the most common motivations behind the use of nonparametric regression over standard parametric methods. Among them is the ability

of nonparametric methods to model complex, nonlinear relationships in a variety of situations. These models do not have to conform to a certain shape or pattern, which is thought to be very true of many real-world processes even though parametric models are most often used to describe them. This is thought to give better predictions with less bias for data where the underlying function is truly nonparametric, an idea that is difficult to dispute.

Another highlighted motivation is the flexible treatment of outliers with nonparametric methods. While there are some modern techniques for dealing with outliers in parametric methods these days, such as methods with high breakdown points [5], the point can also be made that when using parametric methods that the reason some observations are labeled as outliers is not that they don't fit in the pattern of the data, but they don't follow the prescribed parametric distribution. This is not the case with nonparametric regression, and these methods offer a variety of down-weighting techniques that are a good alternative to many of the parametric techniques that entirely ignore or throw out the outliers.

Nonparametric regression also enables the estimation of derivatives, where this estimation can be limited in parametric regression. For example, 2^{nd} degree polynomial regression provides a strictly constant second derivative, while linear regression does not provide a second derivative at all. While some fields of interest have no use or need for a higher order derivative there are some areas where it becomes quite essential. A very popular example of where the second derivative is very useful in a real-world situation is in human growth data, as discussed in Ramsay [6] chapter 6, and Charnigo (2011) [7].

Although parametric and nonparametric statistics are often at odds, using them together can be an invaluable resource that is not taken advantage of often enough. There are countless ways the two fields can be used to complement each other. For example, nonparametrics can be used as a tool to confirm your belief that the use of a certain degree polynomial is appropriate for the data.

Challenges

As with many things in life, some of our strengths can also be viewed as weaknesses, and nonparametric regression is no different. Since we are deriving both the structure and estimates for the model from our data, we require a larger sample size to do nonparametric regression, which is not always possible or desirable. Putting so much emphasis on the data can lead to overfitting in different ways than we often see in parametric statistics. When fitting these nonparametric models, it is easy to allow the training data to control the fit too much, giving us higher variance and lower bias in the training set, and if we let this happen our models can be less effective when we have testing data or new observations. See Mroz and Savage (1999) [8] for a detailed discussion on overfitting.

One important distinction between parametric and nonparametric statistics is the use of parameters. In parametric statistics a parameter is a defining characteristic of a certain distribution that we can pull from our data, like the mean and standard deviation. Most families of distributions only require a couple parameters to describe them in their entirety. This is quite different from the tuning parameters we find in nonparametric regression. Here, tuning parameters are values used to specify the conditional mean function that have to be chosen in order to fit the nonparametric model. As we will discuss deeper later, there is no one correct value of a tuning parameter and their selection can be considered one of nonparametric regressions greatest weaknesses.

Nonparametric statistics is no stranger to the curse of dimensionality [9], in fact nonparametrics may be even more impacted by the curse than parametric statistics. For this reason, it is often the case that semi-parametric methods be explored, such as in Martins-Filho and Yao (2011) [10], in order to try to balance the strengths and weaknesses of each field, though we will not be discussing semi-parametric methods in this paper [11]. Additive models (where interactions are not allowed) have also been used to mitigate the effect of the curse of dimensionality but making the assumptions that there are no interactions in your model should not be taken lightly.

Local Regression

The first nonparametric regression method we will explore is local regression, which has a fairly intuitive name (often a rare occurrence in this field). Local regression originated in the 70's [12] and was an expansion of kernel regression, the second method we will discuss in this chapter. The goal of local regression is to minimize a locally weighted sum of squares, usually over a grid or set of x values that span the support. The weights vary depending on proximity of the input value from the subject, x_i, to the grid value, x_0, for most choices of weights. The closer x_i is to x_0 the higher the weight and vice versa. This is quite different from ordinary least squares regression (OLS) where the goal is to minimize the global sum of squares one time. With local regression we locally estimate μ and μ'. With this method we can essentially "connect the dots" to get a global estimate, but these estimators are not self-consistent, a topic we will dive deeper into later [7].

In 1979 Cleveland improved upon the standard local regression by introducing his robust locally weighted regression as a way to reduce the influence of outliers even more [13]. The first step in Cleveland's procedure is standard local regression. What makes this procedure more robust over the standard is the reassignment of weights based on a point's residual in the standard local regression step. Once the new weights are determined we fit another local regression using them and the cycle continues until it is terminated, or a convergence criterion is met. Just as before, the farther a point is from x_0 the smaller its weight will generally be, which makes good intuitive sense. This is a kind of supervised learning technique and is less sensitive to fluctuations in the training data than standard local regression. See Hastie et al Ch. 2 [14] for more information on supervised learning.

The range around x_0 used for each local estimation is also important because, although they will inevitably overlap, the larger the range the smoother the overall estimation will end up, potentially producing high amounts of bias. Alternatively, smaller ranges lead to overall estimates

that are more tailored to the training data which can lead to higher amounts of variance. Choosing the range, as with all other tuning parameters, is a balancing act and should be treated with care. It is important to remember that this method only produces a set of smoothed points. These points are often connected for a visual representation of the pattern along the data, which can be used advantageously but also misinterpreted easily. For example, there should not be too much meaning assigned to small artifacts in the curve and unfortunately it often requires over-smoothing to get rid of them, so there needs to be an understanding of what the curve represents.

Though local regression is highly touted there are quite a lot of tuning parameters that need to be set, three and a weight function, but Cleveland provides much guidance on their selection. There are four fairly intuitive requirements for the weight function and the author suggests that the tricube weight function can be used in nearly all situations. Among the other parameters that must be set is the order of the polynomial used for the local regression. Here the author suggests that using degree 1 gives adequate fit while remaining computationally unburdensome. Another parameter to set is the number of iterations to carry out. Again, the author suggests 2 iterations is enough based on vast experience. While I am willing to accept a degree 1 polynomial without much discomfort, I think given the computation power of popular programs like R these days it would not be difficult to define a convergence criterion for the continuation of these iterations. The final and most complicated parameter that must be chosen is used to determine the level of smoothing that is to be done. This is where the dreaded subjectivity of smoothing methods comes into play. The smoothing parameter is much more sensitive to the data at hand and must be chosen carefully. With its value ranging from 0 to 1 it is a balancing act between minimizing variability of the smoothed points and staying true to the pattern in the data. This is not to suggest that 0.5 is the optimal choice to balance the two. All data sets are unique and different values of the smoothing parameter will be appropriate. Though the suggestions for this parameter are much softer than previous ones, the author suggests choosing a value between 0.2 and 0.8 for most data and starting with a value of 0.5 when there is less of an idea of what to use.

Plotting a smoothed curve through a scatter plot can be beneficial for seeing general patterns in the data but it is very easy to over-fit the line and the limits of such overfitting are quite subjective. One of our biggest assets in evaluating appropriateness of fit, as supported by Loader [15], is various plots of our data and their residuals. As Loader suggests, there are many ways to detect an inadequate fit to your data in the direction of undersmoothing, but once you achieve a good fit it can be difficult to use residuals to assess if you've gone too far and overfit the data.

Local regression can be a good method to visualize a smoothed curve through a 2-dimensional scatterplot, but most real-world problems involve many more than 2 dimensions. Though having outliers is certainly not the only reason to use nonparametric regression, both local regression and its robust counterparts seem to be a good way to handle fitting a nonparametric curve through the data if you have reason to believe you do not want your fit largely influenced by them. But what happens if the outliers in your data are genuine data points? Should you really be down-weighting them if you have no reason to believe that these points are not in some way truly representative of the pattern in the data at hand? Maybe not, but then again, one can view this method as a way to model the data in a desirable way without completely throwing out valid data points (which is often frowned upon). Nonparametric regression is also not the only solution to such problems. There are many good parametric techniques that have been designed to be robust to outliers in the data, like least trimmed difference [16] and least median of squares [17]. This brief discussion is here to serve as a reminder that there are many things you must consider when modeling in the real world, and not all of them deal with the goodness of fit.

Kernel Regression

Kernel regression, which originated in the 60's [18,19], gave rise to the previously mentioned local regression and lives in the realm of local location estimators along with local regression and nearest neighbors, which will be briefly discussed at the end of this section [20].

While some could argue that kernel methods were improved with local regression it is still a common nonparametric regression procedure, in part because it is easily useful for the proving of theoretical results [21]. In essence, kernel regression takes a weighted average of y_i observations whose x_is are near every x_0 observation in the training data set to obtain an estimate of the response function. The nature of the weighted averages is determined by choice of kernel function, of which there are many common choices including Gaussian, Epanechnikov [22], and tricube (see [4] sections 3.1 and 4.5). It is usually nonnegative and symmetric with a peak at 0. When choosing a kernel function, it is often more natural to pick one with an infinite support set, like the Gaussian kernel, because we would not want to throw away any real-world data. On the other hand, choosing a kernel with a compact support set can be much easier in terms of theory and computation. When using a kernel with compact support you can easily end up with a response function that is not smooth. For example, when using a uniform (-1, 1) kernel the response function will resemble a step function. This may be appropriate in some cases but not others. Kernel selection is often subjective, but some procedures have been proposed to make it less so, as in Ding and Liao (2017) [23].

While kernel selection is very important it turns out that the estimated response function is much more sensitive to the selection of the smoothness tuning parameter *h* [4 section 4.5], which is called the bandwidth. The bias-variance tradeoff is always a hot topic in the modeling world [24] and we can see this directly relates to kernel regression through the bandwidth parameter. The bandwidth parameter *h* must be carefully chosen so that it is not too tailored to the training data set at hand, yet reasonably minimizes the variability (see Hazelton [25] for more on bandwidth selection). Smaller bandwidths lead to noisier estimates while larger bandwidths lead to smoother estimates (*h* is constrained to be a positive value and directly depends on the range of the data).

Evaluating bias when using kernel regression can be difficult because it relies on an asymptotic formula [4 (chapter 4)] which we most of the time do not have enough data to approximate, and also includes an unknown functional of mu. Estimating the variance tends to be

much easier because it does not contain an unknown functional of mu and it is proportional to 1/(n*h*), where n is the sample size and h is the bandwidth. Estimation of the variance does, however, use sigma^2 and often the assumption of homoscedasticity is made, which is not always reasonable [26].

As a way to mitigate the effect of biases sometimes higher order kernels are used. These kernels can be negative, where standard kernels are generally strictly positive, which can lead to some questionable estimates. Using higher order kernels is much more complicated and they are used far less often because substantial reductions in bias are seen only in asymptotically large data sets, which are less uncommon than they were in the past, but still not often used. See Hardle section 4.6 [4] for more discussion on the issue. As an alternative, local regression is often used over higher order kernels as a solution to unwanted bias, though it is important to note that local regression does not entirely eliminate it.

Kernel regression, in addition to being expanded upon with local regression, has also been adapted to what is now known as nearest neighbor's analysis [27]. In kernel regression the bandwidth is specified and held constant for each local value of x_0, which means that a varying number of neighboring points may be used for estimation at each x_0, depending on how dense the region is. Nearest neighbors' analysis takes this idea and adapts it so that instead of fixing the bandwidth and letting the number of neighboring points to vary, the number of neighboring points used for estimation at each value of x_0 is fixed and the bandwidth is allowed to vary to accommodate the specified number of neighboring points.

Self-consistency

Self-consistency is a term coined by Charnigo and Srinivasan (2011) [7] to represent interchangeability of estimation and differentiation. It seems obvious that self-consistency would be a desirable property of most estimates, though this is unfortunately the exception, not the rule,

as we have seen with local regression so far. Some methods like kernel regression and spline smoothing can be made to be self-consistent, but at the cost of much higher bias, which is not desirable. The authors propose a new estimator, the compound estimator, which has many desirable properties. The estimator is formed via a weighted average of many polynomials defined with pointwise estimators and is infinitely differentiable, achieves nearly optimal convergence rates in large samples, and holds self-consistency. It is important to note here that the compound estimator remains infinitely differentiable and self-consistent regardless of the method used for the pointwise estimation. A reasonable method for pointwise estimation is required for probabilistic consistency to hold.

Compound Estimation

Compound estimation should be considered as a strong candidate for use in nonparametric regression scenarios, though it is not without issue. While compound estimation out-performed both local regression and smoothing splines in simulations, there was no R package created for convenient use of this method. This largely decreases the frequency of its use because it can be computationally intensive to compute the convolution smoothing tuning parameter and it can be difficult to convince a researcher to use new methods, much less ones that have not been implemented in their preferred software program.

Tuning Parameters

Tuning parameter selection is arguably the most subjective aspect of nonparametric statistics, as we have seen with each method so far. Choosing tuning parameters wisely can make or break any nonparametric analysis [28], so it's very important to clearly state how they are chosen because poor selection can easily discount the validity of your results. That is not to say that there is one correct way to choose your tuning parameter, in fact there is no one correct way to do so.

Selection techniques can range from very subjective selection based on experience, as Cleveland suggested when we discussed robust local regression, to largely not subjective when choosing the parameters strictly based on some criterion, though choosing which criterion to base your decision on can be seen as subjective as well. More often than not it is very difficult to know whether you have chosen appropriate values for your tuning parameters because in the real world we do not have the true mean response function to compare fitted values to, so we have to compare them to the raw data. This is as opposed to simulation studies where we know exactly how the data originated and are able to confirm the best selection of the tuning parameters [29].

There is an endless list of criteria that have been used for tuning parameter selection, but here I will highlight just a few of the most well-known, and perhaps one that should be added to that list. Among the most widely used criterion are generalized cross-validation (GCV) and the Akaike information criterion (AIC). GCV and AIC are common and easy to use as they have been used for quite some time and implemented in many software programs like SAS and R. However, they are not without their shortcomings. Both of these methods tend to produce highly variable estimates of smoothing parameters that typically lead to under-smoothing [30]. Arguably one of the biggest problems in nonparametric statistics is the continued use of general methods of tuning parameter selection, like AIC (which originated for parametric models), that do not account for the complexity of the situation. Hurvich et al [30] proposed an improvement to AIC for the selection of smoothing parameters in nonparametric models, AIC_c. They showed that their corrected criterion was less biased than the standard AIC while also producing estimates of the smoothing parameter that did not tend to under-smooth and it has been referenced even in recent work on tuning parameter selection for use over the classic AIC [31]. Another very common criterion for selection is Mallows C_p criterion, which gives estimates very similar to GCV [31].

The vast majority of tuning parameter selection criteria have been developed in regard to the mean response function, but in 2011 Charnigo et al developed the generalized C_p criterion (GC_p) for tuning parameter selection with a particular focus on derivative estimation that can be used with

multiple nonparametric regression methods [32]. It was developed to emphasize the estimation of the derivative by defining a proxy for error sum of squares in estimating the derivative mu-prime(x). This criterion can be used with many of the most popular nonparametric regression methods, like kernel smoothing, local regression, and smoothing splines. The main requirement of this criterion is that at any fixed value of the tuning parameter the estimated derivative should have a linear representation in terms of the observed outcomes. This also makes bias and variance calculations easier. Once the tuning parameter is selected the estimator is no longer linear because a nonlinear relationship exists between both the tuning parameter and outcome and the model fit and the tuning parameter. See Charnigo et al 2011 for much more detail and theoretical results [32]. This criterion was tested against a multitude of other selection criteria and out-performed nearly all of them in a variety of settings. This criterion was later expanded to a multivariate GCp (MGCp) for use with multiple covariates (Charnigo and Srinivasan 2015) which appeared in simulations to not be prone to undersmoothing like many of the traditional criteria based on the mean response function.

Conclusion

As we have seen, there are many things to take into consideration when deciding whether or not to use nonparametric regression methods. Often in research there is some form of hypothesis test desired to test the presence of a relationship. Parametric statistics are well suited for this, but nonparametric statistics is often better suited for characterizing a relationship as opposed to testing for one. While it is very common for professionals from other fields to perform some statistical procedures, they must be careful and attentive to detail if they try the nonparametric route. It is important for statisticians to accurately convey the complexity of these models to others, as the theory of nonparametrics can often be quite dense [33,34].

Aside from the careful work a nonparametric method requires, the statistician carrying out the methods must also consider their audience and whether or not they will be pleased with and/or

understand the results of a nonparametric model. All of this is to say that parametric and nonparametric statistics both have their place, and we should be careful to use the best method for the task at hand.

Mixture Modeling Review

Introduction

Mixture modeling can be thought of in more than one context in statistical analyses. In one case we can think of the mixture density estimation problem where you know that your population density is a mixture of a certain number of sub-populations (with certain mixing proportions), but information on the sub-populations is not available for the observations in the sample, similar to a situation where we have an unobserved group indicator that we know exists even if we can't identify it. In this situation you can use the model to estimate which sub-density an observation belongs to if the variable on which the density is split is not available. The second, less common context in which we think of mixture modeling is when you have access to all the variables you need, including a variable on which you believe there may be multiple sub-populations. In this situation you would use the information available to you to try and determine if there is a split in the density in one of the variables that you have recorded. Having access to the variable on which there are yet undetermined sub-populations is an ideal situation, but unfortunately, we often are not able to identify said variable, much less have recorded information on it.

True mixture distributions are often disguised as unimodal distributions. You cannot always see a true multicomponent distribution behind a unimodal plot, even if we know sub-populations are present. It can be hard to see sub-groups separations if the means of the components we do have information on are very similar. Additionally, data visualizations get harder with more and more dimensions involved in the distributions.

For the purposes of real-world applications of mixture modeling it is important to note that we have to restrict our analyses to mixtures containing a finite number of components. Finite mixture densities are densities in which the population of interest is a mixture of J sub-populations (or component populations) with certain mixing proportions [35]. There are many mixture densities containing possibly infinite numbers of components, but we focus on mixtures with relatively few

components for practical applications. The marginal distribution of the outcome Y is a finite component mixture, while the conditional distributions give the components of the marginal distribution.

Parameter Estimation

We will refer to J as the parameter for the number of components within the mixture distribution. Multiple parameters need to be estimated in mixture modeling, including the number of components, J, and their mixing proportions. Parameter estimation is relatively simple if you know J and observe x and y because you can use expectation maximization (EM). Discussed in more detail later, EM is intended to give an approximation of maximum likelihood estimation for the parameters as each iteration increases the likelihood. Even if we don't end at the correct answer via expectation maximization, it will be better than the initial value. Initial values for EM are very important because poor choices can lead to needing a large number of iterations before convergence or getting stuck at a local maximum instead of being able to make it to the global maximum. Although EM is common, there are many other ways to estimate the parameters needed in mixture modeling.

Evolution of Mixture Modeling

We closely reviewed [35] for the history of mixture modeling methods. One of the first methods proposed for solving a mixture density estimation problem where all observations are unlabeled was the method of moments presented by Pearson [36]. This method was quite computationally intensive before computers were of popular use, as it involves lengthy algebraic manipulation of sample moments even in simple cases. For that reason, the method was used sparingly and generally for the simplest case of a mixture of two normal densities. The method of

moments has since been expanded upon and has the desirable aspect that the moment equations are linear in the mixture proportions.

Perhaps more popular than the method of moments for estimating parameters that determine a mixture density was maximum likelihood estimation once it came about in the 1960's [37,38], and it is arguably still the most common method used. Maximum likelihood estimates for the parameters that determine a mixture density will find the parameters that maximize the induced density function of a given sample of observations. With computational burden greatly decreased due to use of computers, maximum likelihood estimation paved the way to increasingly more complicated mixture solutions involving any number of Normal distributions, mixtures of multivariate Normal distributions, samples with labeled and unlabeled observations, and eventually solutions involving mixtures of non-Normal distributions. The maximum likelihood approach involves obtaining a set of likelihood equations and solving for the maximum likelihood estimate. This process is described in much detail by Redner and Walker [35].

Even given its popularity, there are many criticisms of maximum likelihood estimation in this setting [39]. In spite of these inevitable downfalls, maximum likelihood estimates perform well when compared with other mixture density estimation methods [35].

All the talk of method of moments and maximum likelihood is not to say that there weren't other methods developed and/or frequently used in mixture modeling estimation. There were countless methods developed for solving mixture densities and they covered very general [40,41] to very specific [42] cases, with varying levels of success and applicability. However, the method of moments and maximum likelihood estimation represent major milestones in the evolution of mixture modeling methods.

Even with the great popularity of maximum likelihood estimation, these estimates can be difficult to obtain. Computational problems can arise easily because of the "complex dependence of the likelihood function on the parameters to be estimated." [35] With mixture density problems

it becomes difficult to find solutions because after the (log) likelihood function has been differentiated it is not linear and you must use an iterative procedure to approximate the solution.

Many iterative procedures have been used for mixture density solutions, but one method that has been widely accepted and has many desirable properties is the expectation maximization (EM) algorithm [43]. It would almost be inappropriate to attribute the formulation of the EM algorithm to one set of authors because it was seemingly independently derived by multiple authors [44-47]. The common theme among developments of the EM algorithm is equating partial derivatives of the log likelihood to zero to obtain equations for the algorithm. It appears to be common practice for methods in the mixture density field to originate for simple cases of a mixture of a couple univariate Normal distributions and then be extended to more complicated scenarios, and the EM algorithm was no different. Although, most applications of EM have involved mixtures of densities belonging to the exponential family of distributions because they are particularly easy to implement.

A different viewpoint of the EM algorithm was taken by Dempster et al [43]. Instead of building on partial derivatives of the log likelihood, Dempster et al approached mixture densities as "an estimation problem involving incomplete data by regarding an unlabeled observation on the mixture as "missing" a label indicating its component population of origin," which seems like a very intuitive approach. With this formulation they showed that the mixture density setting of the EM algorithm was truly a special case of a more general EM algorithm which can be used in a wide variety of settings for "approximating maximum-likelihood estimates from incomplete data," which is a very useful result. Redner and Walker [35] believe this is the best conceptualization of the EM algorithm in this setting.

The iterative procedure of the EM algorithm involves an E-step and an M-step, unsurprisingly standing for Expectation and Maximization, which have both been made relatively easy with modern computation power. With each iteration the log-likelihood function increases (monotonically) and the algorithm relatively reliably converges to at least a local maximum for the

log-likelihood. With an appropriate initial starting point the algorithm can be assumed to find the global maximum. It is admittedly a slight injustice to summarize the entire EM algorithm to just a few sentences, so please see Redner and Walker [35] and Dempster et al [43] for much more detail on the general EM algorithm and how it specializes to the mixture density estimation problem.

Although some would argue EM is the best method for mixture density estimation, the EM algorithm is not without fault. In many applications the convergence of the iterations can be very slow, though this becomes less and less of an issue as modern computing power continues to grow. As one would expect, increasingly poorly separated mixtures will take an increasingly large number of iterations to approximate, but promising application results have been presented by Redner and Walker [35] and others that suggest the EM algorithm still performs well in these cases after a relatively low number of iterations.

Redner and Walker [35] also touch on two important topics for the feasibility of any mixture density estimation problem: identifiability and information. The Fisher information matrix can provide information on how good you can expect your estimates to be, while identifiability refers to the possibility of unique parameter estimates. See section 2.5 of [35] for an explanation and importance of these concepts, as well as many references for a deeper understanding.

Likelihood Ratio Testing

A popular way of determining if your data comes from a mixture distribution is by using a likelihood ratio test (LRT). We can test the null hypothesis that the distribution has p components versus the alternative that it has q components, where p and q are integers, $p<q$, and p can equal 1. The test is not restricted to testing consecutive numbers of components, though it makes good sense to do so. If multiple tests are performed on consecutive numbers of components, care must be taken in the consideration of the usual multiple testing precautions. Using a Bonferroni correction will likely be conservative, and the tests will not be independent.

A typical likelihood ratio test is easy to employ and has a chi-squared limiting distribution (see Statistical Inference [48] section 10.3). In the mixture modeling setting unfortunately not all of the regularity conditions for the test are met. A mixture of two normal distributions can be written as $(1-\alpha)N(\theta_1,1) + \alpha N(\theta_2,1)$ and a test of homogeneity (for 1 vs 2 components) has a null of H_0: $\alpha(1-\alpha)(\theta_2-\theta_1)=0$. We can see here that there are multiple ways that this null hypothesis can be true. Either α can be 0, $1-\alpha$ can be 0, or θ_2 can equal θ_1, the latter of which is the true homogeneity result that we want to test. The null hypothesis here lies on the boundary of the parameter space and if it is true then the parameters are not identifiable [49]. We also require a positive, finite Fisher information number here, which can fail and coupled with loss of identifiability contributes to the more complicated nature of the LRT in the mixture modeling setting.

The limiting null distribution of the LRT in mixture modeling is much more complicated than a chi-squared distribution because of the violations in the regularity conditions, two of which were mentioned above. As shown by Chen and Chen [50], under a new set of conditions, one of which is a compact parameter space, it turns out that the limiting distribution of the likelihood ratio test statistic is "the squared supremum of a truncated standard Gaussian process". The supremum is taken over the parameter space and removes the separation condition of Ghosh and Sen [49] that the parameters are not identifiable, as described above.

This is clearly a rather complicated distribution to pull a quantile function from in order to get an asymptotic p-value. Though there have been multiple methods proposed for obtaining a p-value [49], Chen and Chen recommend using a bootstrap procedure, which they accredit being based on Beran (1988) [51]. The authors condone this procedure because it requires no specification of the null distribution of the data and no "detailed knowledge of the limiting null distribution of the LRT," merely that it exists [50].

Beran was far from the only one to employ bootstrapping for p-values in the mixture modeling LRT setting. McLachlan (1987) [52] discussed the method in the restricted case of a mixture of one versus two univariate normal distributions (with a common variance), a theme we

have already identified as common throughout the history of mixture modeling methods. Under this bootstrapping procedure, which is inherently parametric, a very large number of independent samples is taken from the distribution formed under the null hypothesis, in this case a normal distribution with mean and variance corresponding to the original sample of data. The value of the test statistic (-2log-likelihood) is computed for each bootstrapped sample and the distribution of those many values is used to approximate the true null distribution of the test statistic. Finally, we can use the quantiles of this distribution to obtain approximated p-values for the hypothesis test given the value of the test statistic from our original sample. Through simulation McLachlan showed that the null distribution of the -2log-likelihood is very similar to that of a chi-squared distribution with 2 df and therefore appropriate for the approximation of p-values when n is large enough (>100), but this only holds when the variances of the components in the mixture are the same [52].

There are many well-knows problems that come with using bootstrap procedures. In the situation of a parametric bootstrap, like the one described above, the performance of the test can greatly rely on the specification of an appropriate null distribution. The sample size and representativeness of the original sample can affect the quality of the parameters used to specify the distribution that the bootstrap samples are drawn from. Further, the number of bootstrap samples taken from the proposed distribution can have a great influence on the formulation of the approximated distribution of the test statistic. If the number of bootstrap samples taken is small then the variance of the quantiles of the distribution will be high, but it can sometimes take a considerable amount of time to bootstrap a large number of samples. In the real world the best we can do is assume that we have correctly specified our distributions and taken enough bootstrap samples. Bootstrapping is a very useful tool, but we would not recommend heavily relying on it to obtain p-values when there are arguably more reliable methods available.

Modified Likelihood Ratio Testing

A modified likelihood ratio test (MLRT) was presented by Chen, Chen, and Kalbfleisch (2001) [53] as a way to try to mitigate the violation of identifiability in the standard LRT and identify a simpler limiting distribution to gain p-values from with minimal loss of power. The MLRT places a penalty on the log-likelihood function so that the mixture weight α is pushed away from 0 and 1 (the boundaries of the parameter space) with the goal that the only way to retain the null is when θ_1 and θ_2 are equal. This reduces the non-identifiability of the parameters. With this penalty comes the burden of appropriately choosing the value of this additional penalty parameter, C, which must be a fixed positive constant. Chen, Chen, and Kalbfeisch [53] suggest that C=log(M) is a suitable choice when the parameter θ if the kernel density is restricted to values of (-M, M), as their simulations showed that the MLRT is not sensitive to choice of C. This choice of penalty can be modified in the situation that the parameter space of a certain kernel does not allow negative values.

The penalty placed on the log likelihood results in a much simpler limiting null distribution that is asymptotically most powerful under local alternatives, meaning it has the most power to handle alternatives that are converging to the null at an appropriate rate. This method is only proposed for when the parameter θ is one-dimensional. The limiting null distribution is an equally weighted mixture of a central chi-square with 1 df and a central chi-square with 0 df. Note that a chi-square with 0 df is a degenerate distribution with a point mass at 0. In this setting the homogeneous population pdf is referred to as the kernel function, which must meet certain criteria for the MLRT limiting distribution to hold [53]. Through simulations the authors showed that the MLRT performed as well or better than other methods such as Neyman and Scott's C(α) test [54,55], Davie's method [56] and the bootstrap LRT [52] in a variety of situations, but some common kernels, such as kernels from the exponential family, do not meet the criteria needed, which can be a major limitation.

D-Testing

Taking an entirely different approach than the LRT and MLRT procedures, the D-test of Charnigo and Sun (2004) [57] can also be used to test homogeneity in mixture modeling. The D-test uses maximum likelihood estimation to obtain estimates of the parameters of the mixture components belonging to a certain parametric family of distributions. Charnigo and Sun point out that the selection of a specific distribution in the alternative gives the test the most power when that alternative is correct, but it can be generalized to a nonparametric alternative at the loss of some power if desired. They also recognize that any reasonable parameter estimation method other than maximum likelihood estimation can also be used here.

The D-test statistic is the integrated square of the L^2 distance between a fitted homogeneous and a fitted heterogeneous model, which will be small when the mixture is homogeneous and large when it is heterogeneous. The statistic uses only the estimated parameters of the distributions, and therefore changes depending on the specified family of distributions of the mixture components. The D-test consistently outperformed the MLRT in simulations with mixture components from a normal location family and when the sample size is large [57] The D-test can better detect differing shapes (rather than scales) between mixture components because it is based off the L^2 distance.

When the sample size is not large, or the mixture components come from a family other than the normal location family, Charnigo and Sun [57] suggest using a weighted D-test. The only difference with the weighted D-test is that it adds a weighting function that is designed to make it easier to identify separations in the components with the L^2 distance. Using a weighting function can be thought of as analogous to placing a transformation on the data before computing the D-test statistic. Different transformations of the data will result in different weighting functions and when chosen appropriately the weighted D-test performs favorably to the MLRT and much better than the standard D-test.

One perceived advantage the D-test has over the MLRT is that the test statistic depends on the data only through the parameter estimates, and therefore once the parameters are estimated the

rest of the data can be disregarded for the remainder of the procedure. This is contrary to the MLRT which uses all of the data to calculate the test statistic. With relatively small datasets this will not seem to matter much, but with "big data" becoming more and more popular and accessible this trait seems more desirable. Not only does it greatly cut down on computation time of the test statistic, but data storage issues do not present a problem for this test.

While there are definitely advantages to using the D-test we need to consider the implications of a much stricter alternative than the MLRT. The inherent parametric assumptions of the D-test may not always be reasonable or desirable. In some cases, the more general alternative of the MLRT will be preferred.

Conclusion

Here we briefly covered some highlights in finite mixture modeling methodology, but existing methods reach far beyond this review and are expanding every year. Two major milestones in mixture modeling were the development of method of moments and its subsequent replacement with maximum likelihood estimation as the most popular method of the time. Log likelihoods became a vital tool for mixture modeling methodology through their use in MLE, EM algorithms, and LRTs. MLRTs improved upon LRTs by dropping bootstrapping for p-values and implementing a penalty on the log likelihood to simplify the limiting null distribution, resulting in a test that outperformed those previously used. The D-test, however, approaches mixture modeling from a different direction. The D-test statistic is based on the L^2 distance between a fitted homogeneous and a fitted heterogeneous model. This method has outperformed MLRT in simulation and can better detect differing shapes between mixture components because of its basis in the L^2 distance.

Computational considerations, which were once considered a heavy burden in mixture modeling, have been greatly reduced with computational advancements and statistical programs. Many mixture modeling techniques were developed for the easiest case, a mixture of two Normal

distributions. Increases in computational power have facilitated the use of mixture modeling for situations beyond a mixture of 2 distributions and expanded its horizons to include non-Normal distributions as well.

CHAPTER 2: Associations of novel plasma-based biomarkers of neurodegeneration, angiogenesis, and inflammation with demographic and clinical characteristics among cognitively normal research volunteers

Taylor G. Estepp, MS; Richard J. Charnigo, PhD; Erin L. Abner, PhD; Gregory A. Jicha, MD, PhD; Tiffany L. Sudduth; David W. Fardo, PhD; Donna M. Wilcock, PhD.

Abstract

This study examined the relationships between 13 novel blood-plasma biomarkers and dementia-related demographic and health factors in a cohort of 237 cognitively normal research volunteers. We regressed each biomarker on selected covariates to explore the relationships the biomarkers have with health factors likely along the causal pathway to dementia to assess whether these factors may contribute to biomarker values. In this sample, biomarker concentrations were largely not associated with participants' demographics or health conditions, but some expected associations (e.g., of APOE with Aβ42/Aβ40) were observed. To assess robustness of cross-sectional associations, we used a second set of measures obtained five years later on participants who remained cognitively normal and found similar results.

Introduction

By 2030, all baby boomers – approximately 21% of the US population [58]—will have reached 65 years of age [59]. As the aging population grows, so does the prevalence of age-related conditions. According to the Alzheimer's Association, in 2020 more than 5 million Americans were living with Alzheimer's disease (AD) dementia, and 1 in 3 deaths among adults aged 65 and older were associated with dementia [60]. Dementia is characterized by cognitive impairment that affects memory and other cognitive functions (such as the ability to reason, plan, or effectively communicate) [61] and is most prevalent among adults over age 65 [62]. Both the prevalence of dementia and the accompanying burden on patients, caregivers, and health care systems are expected to grow in the coming years, absent significant headway in treatment and prevention measures.

Decades of research have characterized the pathophysiology and natural history of the diseases that cause dementia, with a heavy focus on AD [61]. However, both dementia and AD remain incurable and without proven prevention measures [62]. Moreover, diseases that cause dementia are complicated [63], and the gold-standard diagnosis for AD still requires brain autopsy. Clinical diagnosis of AD is based on symptoms and patient history, and may also include CSF (e.g., for amyloid and tau) [64] and neuroimaging biomarkers [65]. Neuroimaging can be a very expensive procedure, and attaining CSF is considered an invasive procedure [66]. These, along with other reasons, make CSF and neuroimaging biomarkers not feasible for many studies [67]. Given such limitations, there is an urgent need for valid, accurate, and easily measurable biomarkers of dementia-causing diseases.

Blood-based biomarkers for AD and related dementias (ADRD) are a recent development and have not been thoroughly studied or optimized. However, there is hope that blood-based biomarkers can help identify disease states and predict cognitive trajectories, while mitigating difficulties associated with more invasive and expensive testing [66,67]. Recent advancements in technology used to measure biomarkers have allowed us to begin exploring these possibilities. Quanterix Single Molecule Array (SiMOA) technology, which is much more sensitive than previous blood assays, allows measurement of AD-relevant biomarker concentrations in the blood, which are much lower than concentrations in CSF [68].

To inform future studies based on SiMOA data, the current study investigated the relationships between three types of biomarkers (inflammatory, vascular, and neurodegenerative; defined below) and participant demographics, health related factors, and technological factors related to the assays. Our objective was to better understand which, if any, features outside of neurodegenerative or cerebrovascular disease may influence biomarker values [69]. In participants with initially normal cognition, cross-sectional associations were estimated at two time points, five years apart, to assess robustness of these associations. As part of assessing associations between

participant characteristics and biomarker values, we also quantified the proportion of biomarker variance explained by covariates.

Methods

Setting

Data for the current study were drawn from the community-based longitudinal cohort of brain aging and cognition at the University of Kentucky Alzheimer's Disease Research Center (UKADRC). Cohort recruitment began in 1989, and over 1000 participants have been recruited into the cohort and agreed to be followed approximately annually until death; most participants also consent to brain donation [70]. To be included in the cohort, an individual has to live close enough for a brain autopsy to be performed at UKADRC within 4 hours of death. In 2019, UKADRC added a Biomarker Core, which uses SiMOA to measure a standard set of biomarkers in plasma donated by UKADRC participants at their annual visits [70]. Sampling banked plasma for biomarker analysis began with participants who had two visits five years apart, beginning with visits that took place in 2012 and in 2017, as plasma collection at UKADRC was transitioned to Ethylenediaminetetraacetic acid (EDTA; used in the present analysis to standardize samples) from heparinized vacutainer tubes in 2012. The UK Institutional Review Board (IRB) approved all study procedures, and all participants provided written informed consent.

Study Design

We conducted a retrospective study on a subset of the University of Kentucky Alzheimer's Disease Research Center (UKADRC) longitudinal cohort [70]. Biomarker data in the current study were obtained on two pairs of visits: 2012 and 2017 (12/17), and 2013 and 2018 (13/18). If participants were included in both sampling pairs, the data from the 12/17 set were used. Inclusion criteria for the current study were enrollment in the UKADRC cohort, available blood-based

biomarker data in the identified visit pair, and a diagnosis of normal cognition at the first of these visits.

Plasma Biomarkers

The SiMOA-based measures included amyloid-beta$_{1-40}$ (Aβ40), amyloid-beta$_{1-42}$ (Aβ42), total tau (distinguished from other tau measurements such as p-tau), neurofilament light chain (NfLight), tumor-necrosis factor-alpha (TNFα), interleukin 6 (IL6), interleukin 8 (IL8), interleukin 10 (IL10), interleukin 1Beta (IL1B), matrix metallopeptidase 9 (MMP9), and placental growth factor (PlGF). In addition to these individual biomarkers, we also investigated the ratios of Aβ42/Aβ40 and tau/Aβ42 [Table 1]. Notably missing from our biomarkers is phosphorylated tau, or p-tau. P-tau values were not included in this analysis, as the assay used for the 12/17 samples (p-tau 231) demonstrated poor reliability in validation testing (data not shown).

We classified the biomarkers into clinically relevant subgroups: (1) nonspecific neurodegenerative and AD markers, containing NfLight, tau, Aβ40, Aβ42, and the ratios Aβ42/Aβ40, and tau/Aβ42; (2) vascular markers, containing PlGF and MMP9; and (3) inflammatory markers (i.e., cytokines), containing TNFα, IL6, IL8, IL10, and IL1B.

These biomarker data were collected at a single facility using standard NIA/NACC biospecimen best practice protocols, and the biomarker assays were run in the single UKADRC biomarker core biosample laboratory. During the interval between the processing of the 12/17 and 13/18 samples, the machine used to run the assays was updated from the Quanterix HD1 Analyzer to the HDX Analyzer, which incorporated sophisticated control systems to enhance reproducibility and included essential temperature control that the HD1 lacked [71]. It was unclear a priori whether the data produced on the two instruments are interchangeable, especially given the upgraded temperature control feature. Quanterix no longer sells or supports the HD1 Analyzer [72], and we are unaware of any publications comparing the performance of the HD1 and HDX machines. Yet, we assume there are extant HD1 machines in use. Labs that have transitioned from the HD1 to the

HDX, as well as labs considering the transition, may have similar questions about the consistency of their data across machines. Thus, in addition to the covariates described below, all analyses included a batch indicator (12/17 vs 13/18).

Covariate Selection

Because our interest in these biomarkers relates to their potential association with ADRD and vascular cognitive impairment and dementia (VCID), we identified covariates that would mitigate confounding between biomarker levels and measures of cognition, with confounding defined as distortion in the association between the biomarkers and cognition arising from their shared causes. We first created directed acyclic graphs (DAGs) to encode our theoretical model for the causal relationship with cognitive status [73] [Supp. Figure 1-3]. We used the biomarker types (vascular, inflammatory, and neurodegenerative/AD) as the exposure for these DAGs assuming that biomarkers of the same type would have similar causes and effects, rather than generating an independent DAG for each individual biomarker. Covariate selection was guided via sufficient adjustment sets in each DAG [Table 2], which are sets of covariates that theoretically eliminate confounding and bias between the exposure and the outcome when adjusted for.

Participant age (in years), gender, BMI (calculated via measured height and weight), lifetime smoking status (ever vs. never), and *APOE* were included as covariates. *APOE*, the strongest genetic risk factor for late onset AD [74], is included as 0 *e4* alleles vs any *e4* alleles, as our sample size did not warrant a finer categorization. Race was not considered as a covariate, in part because the study sample is primarily Caucasian (>90%), and we lacked the ability to assess racial and ethnic differences in biomarkers within this sample..

Self-reported medical conditions (coded ever vs. never, unless otherwise specified) were cancer, cardiovascular conditions (CV; any vs none), cerebrovascular conditions (CB; any vs none), COPD, depression, diabetes, hypercholesterolemia, hypertension, and vitamin B12 deficiency. CV was operationalized as a single variable to indicate whether a participant had at least one

cardiovascular condition, based on self-reported atrial fibrillation, angina, angioplasty, coronary bypass, congestive heart failure, or heart attack. CB was similarly operationalized based on self-reported ischemic stroke and transient ischemic attack. Medical conditions were updated at annual visits.

Cognition

Participants undergo cognitive testing at each study visit, and these data are used, in combination with results of clinical examinations, to ascertain syndromic cognitive diagnosis: normal cognition, MCI, or dementia [75]. Explicit guidelines for clinical diagnosis were followed at each visit to reduce biases from subjective clinical diagnoses [70].

Statistical Analysis

Biomarkers were investigated individually to estimate their associations with participant characteristics and medical conditions. Adjusted analyses were implemented as 13 linear regression models, with individual biomarkers specified as dependent variables. The first set of analyses focused on the baseline levels of the biomarkers (i.e., first year in the pair 12/17 or 13/18). Biomarker values were log-transformed [76,77] to improve plausibility of linear model assumptions [78]. Log-transformations also present an advantage for analyzing biomarker ratios, in that log-transforming a ratio makes the results invariant to choice of numerator and denominator [79]. Although some values appear as possible outliers [Figure 1, Supp. Figure 4], all values were double checked for errors and confirmed.

In the 13 linear models, the independent variables were based on the DAG sufficient adjustment sets, along with the batch indicator. Missingness in the data was minimal; there were at most 5 missing observations for any covariate, and at most 13% missing observations for any biomarker, though most biomarkers had fewer than 5% missing values. Thus we used only

complete cases, which slightly reduced the effective sample size for each linear model. Goodness-of-fit was assessed via Normal Q-Q plots and plots of residuals vs fitted values. R-squared was used to estimate the proportion of variance explained by covariates.

After analyzing baseline plasma data, we performed sensitivity analysis using information from the second visit (5 years later) among participants who remained cognitively normal. The same covariate sets were used in sensitivity analysis.

Upon reviewing results of the original 13 models, we noticed that the model with the highest R-squared involved a ratio of two individual biomarkers (Aβ42/Aβ40). These two biomarkers had a pairwise correlation of 0.629 [Supp. Table 1], which was much stronger than the correlations between all other pairs of biomarkers, except for TNFα and PlGF, which had a correlation of 0.694. Therefore we pursued a post hoc analysis using log(TNFα/PlGF) as an outcome to see if a linear model for this ratio could produce a similarly large R-squared. Covariate selection for the model of log(Aβ42/Aβ40) had been guided by the sufficient adjustment set for Neuro/AD biomarkers. The same could not be done for the model with log(TNFα/PlGF) because TNFα was classified as inflammatory and PlGF was classified as vascular. Therefore, we used all variables from each sufficient adjustment set (inflammatory and vascular) as covariates in this post hoc model for log(TNFα/PlGF).

We used a 5% significance level when interpreting results, but numerical p-values are supplied in all cases. Analyses were performed with R version 3.6.2 in RStudio, using packages readxl, haven, tidyverse, plyr, ggplot2, and gridExtra [80-86].

Results

A total of 237 initially cognitively normal UKADRC participants met all inclusion criteria. The average baseline age was 82.7 years, 62% of participants were female, and 31% had at least one *APOE* e4 allele [Table 3]. Overall, the selected demographic and clinical features did not

explain a majority of the variance in biomarker values. The mean and median R-squared values were 0.117 and 0.088 [Tables 4-6]. The highest R-squared produced by any of the models was 0.363 for log(Aβ42/Aβ40).

Demographics and Genetics

Of the 13 biomarker variables, 4 were significantly associated with increasing age, which was associated with higher concentrations of Aβ40, NfLight, MMP9, and IL10. For a 5-year increase in age, the models predicted 6.2% (95% CI: 0.5, 13.6), 21.5% (95% CI: 14.3, 29.4), 16.2% (95% CI: 3.0, 31.4), and 11.1% (95% CI: 3.2, 20.1) increases in these biomarkers, respectively. Increasing age was also significantly associated with lower concentrations of Aβ42/40 and tau. For a 5-year increase in age, the models predicted 9.1% (95% CI: 3.5, 13.9) and 8.1% (95% CI: 0.8, 14.8) decreases in Aβ42/40 and tau, respectively.

Gender was not significantly associated with biomarker concentrations in any of the models, and *APOE*, which was only included in the models for the six neurodegenerative biomarkers, was significantly associated with log(Aβ42/Aβ40). This biomarker ratio was predicted to be 18% lower among participants who had at least one *e4* allele (95% CI: 4, 30).

Medical Conditions

Hypertension, included in all 13 models, was significantly positively associated with log(Aβ40) and log(Tau) but not other biomarkers or their ratios. Cancer history, included only in the five inflammatory biomarker models, was significantly positively associated with log(IL6). No other medical conditions had significant associations with the biomarkers.

Batch (HD1 vs HDX)

Eight of the 13 models produced a significant batch coefficient. Effect size was calculated for each batch coefficient and defined as the estimated number of standard deviations that the mean

biomarker concentration changed by in absolute value when the batch changed (i.e., absolute value of each batch coefficient, divided by the marginal standard deviation of the outcome). Our batch effect sizes ranged from 0.06 to 1.17, with an average of 0.56 [Supp. Table 2]. Ten of the 13 batch effect sizes exceeded 1/3 of a standard deviation, and three of the batch effect sizes were greater than 1, meaning that in these cases the biomarker measurements from the HDX were predicted to be more than one standard deviation different from those of the HD1.

Sensitivity Analysis

After 5 years of follow up, most participants remained cognitively normal, while 38 (16%) transitioned to MCI and 9 (4%) to dementia [Supp. Table 7]. We repeated all analyses on the 190 individuals remaining cognitively normal at the second visit (5 years later) [Supp. Tables 3-5]. *APOE* was again only significantly associated with one outcome, though here it was log(Tau/Aβ42). Six of these models produced a significant batch effect, again with the majority showing higher means for batch 1 (i.e., the HD1 yielded higher measurements on average). Otherwise, we saw no discernable patterns.

Post Hoc Analysis

One post hoc model was run for the ratio of log(TNFα/PlGF) based on the high pairwise correlation between TNFα and PlGF. This post hoc model produced an R-squared of 0.098, which was very similar to the average R-squared for the original 13 models [Supp. Table 6]. The only variable with a significant coefficient in this post hoc model was batch (effect size was 0.68 [Supp. Table 2]).

Discussion

We evaluated participant and technical characteristics as predictors of concentrations of novel plasma-based biomarkers in a cohort of cognitively normal research volunteers. Encouragingly, biomarker concentrations were not strongly associated with age or gender, nor medical conditions, suggesting that changes in these biomarkers, when observed, are likely attributable to neuropathological changes rather than other confounding medical conditions. Further supporting this interpretation is the finding of significant relationships of *APOE* with Aβ42/40 and tau/Aβ42. This study adds important data to the literature on plasma biomarkers in cognitively normal individuals given that, of the existing studies reporting on these SiMOA-measured biomarkers, all but one focused on cognitively impaired study participants [87].

The relationships of these plasma biomarkers with demographics of well characterized individuals have not been widely studied. While published plasma biomarker studies have reported relationships of their biomarkers with age, gender, *APOE*, race, education, and/or BMI [68,87-93], the majority of these studies were limited to reporting on only subsets of these factors. Studies using SiMOA technology often only reported these relationships as supporting information, secondary to a main analysis involving the biomarkers [87-90,93], or such information was not reported at all [68,91-92].The breadth of factors we studied in relation to plasma biomarkers stands out in comparison to the current literature. Additionally, the majority of similar studies examined only 4 or 5 potential biomarkers, most commonly Aβ42 and total tau, while we examined a much larger set of biomarkers. The relationships reported are not entirely consistent across the literature, suggesting that study variability in sample acquisition, handling and processing, as well as the methods used for biomarker analysis may influence the results from any given study. Thus, our study also contributes additional information about the influence of demographic, clinical and genetic influences on biomarker measurements. While we were not able to study p-tau because of an outdated assay, its use as a potential biomarker for dementia has been examined and we believe it to have as much potential as any other biomarker we studied here [67,69].

Our sample size of 237 observations in the main analysis, which is is comparable to those reported in the similar literature, is robust in the early era of plasma biomarker discovery in the field of neurodegenerative diseases. Further studies using cohorts with even larger sample sizes should provide clearer insights into the relationships seen in this study. The availability of a longitudinally, well-characterized, cohort with uniform biospecimen collection is a clear strength of the present study. The sensitivity analysis, conducted to assess reproducibility of our results, is also a positive feature of this study.

While our sample was restricted to participants with intact cognition, approximately 20% of participants were subsequently diagnosed with MCI or dementia over the five year follow up period for each cohort in this study. While heterogeneity in our sample is unavoidable, the degree of heterogeneity with regard to preclinical brain disease is likely relatively small and analytically uninfluential given that only a small portion of our participants were diagnosed with some level of cognitive decline within five years of follow-up. We repeated these analyses for the subset who remained normal 5 years later and found no remarkable differences or patterns as compared to the main analysis, suggesting that the individuals who had declined within 5 years had no substantial effect on our results.

While Type II errors (i.e., false negatives) are possible due to sample size, it is also possible that few relationships truly exist. Arguably, our most robust results were for Aβ42/Aβ40. In both the main analysis and the sensitivity analysis, this model had the largest R-squared values, at 0.363 and 0.452 respectively, making log(Aβ42/Aβ40) the most predictable outcome with the most variance explained. Notably, this outcome involves a ratio of biomarkers that is consistent with other reports in the literature [87,90]. The observed relationships with age were not consistent across all biomarkers. There were six statistically significant age coefficient estimates: 4 predicted increases in the biomarkers with advancing age, and 2 predicted decreases. Though not all were statistically significant, 8 of the 13 age coefficient estimates were positive. Other studies examining blood biomarkers using Quanterix SiMOA technology found associations of age with total tau [87-

88,90,93], Aβ42/40 [87], NfLight [87-88], Aβ42 [87,89-90,93], Aβ40 [89,93], and TNFα [90], though the directions of the associations were not always reported. While age was not consistent in direction or magnitude of association with biomarkers in the present study, it may yet influence temporal changes in biomarkers that may be non-linear. Future research will examine temporal changes in biomarkers and cognition, in tandem with age and other covariates.

While the use of the two different Quanterix machines may be regarded as a limitation in this study, these data provide valuable insights into the impact of technologic upgrades within even a single quantitative biomarker assay platform. We note that most SiMOA studies in the comparable literature used the HD1 analyzer [87-93]. The present demonstration of a "batch effect", based on analyzer upgrades, is an important consideration when interpreting results from cross site and even same site longitudinal studies. While inclusion of batch in the linear models compensates for a possible systematic tendency of one machine to give a higher or lower result, its inclusion will not correct distortions that may exist in the associations between covariates and biomarkers, and predictability of some biomarkers could be affected by such distortions.

This study is not without its limitations. Using only complete cases introduces some bias, but the proportion of missing observations is quite modest. Additionally, having all participants in this cohort consent to being followed through autopsy introduces another selection bias, as these individuals are more likely to be highly educated and/or motivated to help with dementia research (perhaps indicating a higher prevalenceof a family history of dementia) [94]. Linear models may be too limited to accurately describe the relationship between biomarkers and covariates as well. We can generalize our results to other cognitively normal individuals in this age range in the state of Kentucky, specifically in the central Kentucky region where the University of Kentucky is located, because our participants comprise a community-based sample from this specific area. However, our results may not be generalizable outside the state, or even to the entire state, because the populations will be qualitatively different in many socioeconomic factors. Kentucky is a high

poverty state, and the outlying rural counties are quite different from the urban region our study sample comes from.

Overall, we present evidence that SiMOA plasma biomarkers do not appear to be strongly associated with medical conditions or demographic characteristics among cognitively normal research participants. However, this does not preclude the possibility that such biomarkers, or temporal changes in them, may still aid in predicting cognitive decline, and our results suggest face validity (e.g. *APOE* was associated with Aβ42/40 but not Nflight). These results encourage the use of plasma biomarkers in future dementia research without considerable worry of confounding due to demographic or medical conditions.

Acknowledgements

We would like to thank the UKADRC research volunteers, their families, and the UK ADRC staff for their part in furthering this research.

Funding

This work was partially supported by NIA grants P30 AG072946 and R01 AG038651.

Tables and Figures

Table 1. Distribution of plasma biomarkers at baseline and five years later among cognitively normal older adult research volunteers

Biomarker Distribution Statistics						
	Baseline			5 years later		
Biomarker	Mean (SD)	Median (IQR)	Effective N	Mean (SD)	Median (IQR)	Effective N
Aβ40	193.59 (104.63)	172 (122.88)	267	237.29 (122.7)	215.12 (130)	266
Aβ42	11.39 (6.97)	10.1 (9.3)	273	14.07 (8.15)	13.25 (10.46)	275
Aβ42/40	0.07 (0.06)	0.05 (0.03)	264	0.06 (0.05)	0.05 (0.03)	264
Tau	6.95 (6.75)	4.94 (3.97)	276	8.46 (47.34)	4.22 (3.38)	277
Tau/Aβ42	0.88 (1.25)	0.52 (0.5)	273	0.64 (3.15)	0.34 (0.22)	275
NfLight	20.61 (23.47)	16.48 (12.89)	276	25.4 (14.49)	21.55 (17.77)	275
PlGF	27.02 (70.19)	4.03 (5.58)	267	26.15 (67.58)	4.29 (5.55)	273
MMP9*	50.72 (66.81)	27.45 (40.45)	273	78.80 (12.59)	38.90 (64.15)	259
IL6	1.61 (4.36)	0.8 (0.91)	275	2.46 (8.14)	0.99 (1.27)	274
IL8	0.41 (1.34)	0.19 (0.33)	250	0.29 (1.3)	0.14 (0.18)	242
IL10	0.7 (1.42)	0.5 (0.35)	274	0.85 (1.69)	0.57 (0.46)	277
IL1b	0.44 (2.63)	0.07 (0.09)	244	0.33 (1.54)	0.02 (0.04)	245
TNFα	1.89 (2.7)	1.21 (0.82)	271	2.69 (9.26)	1.32 (1.07)	266

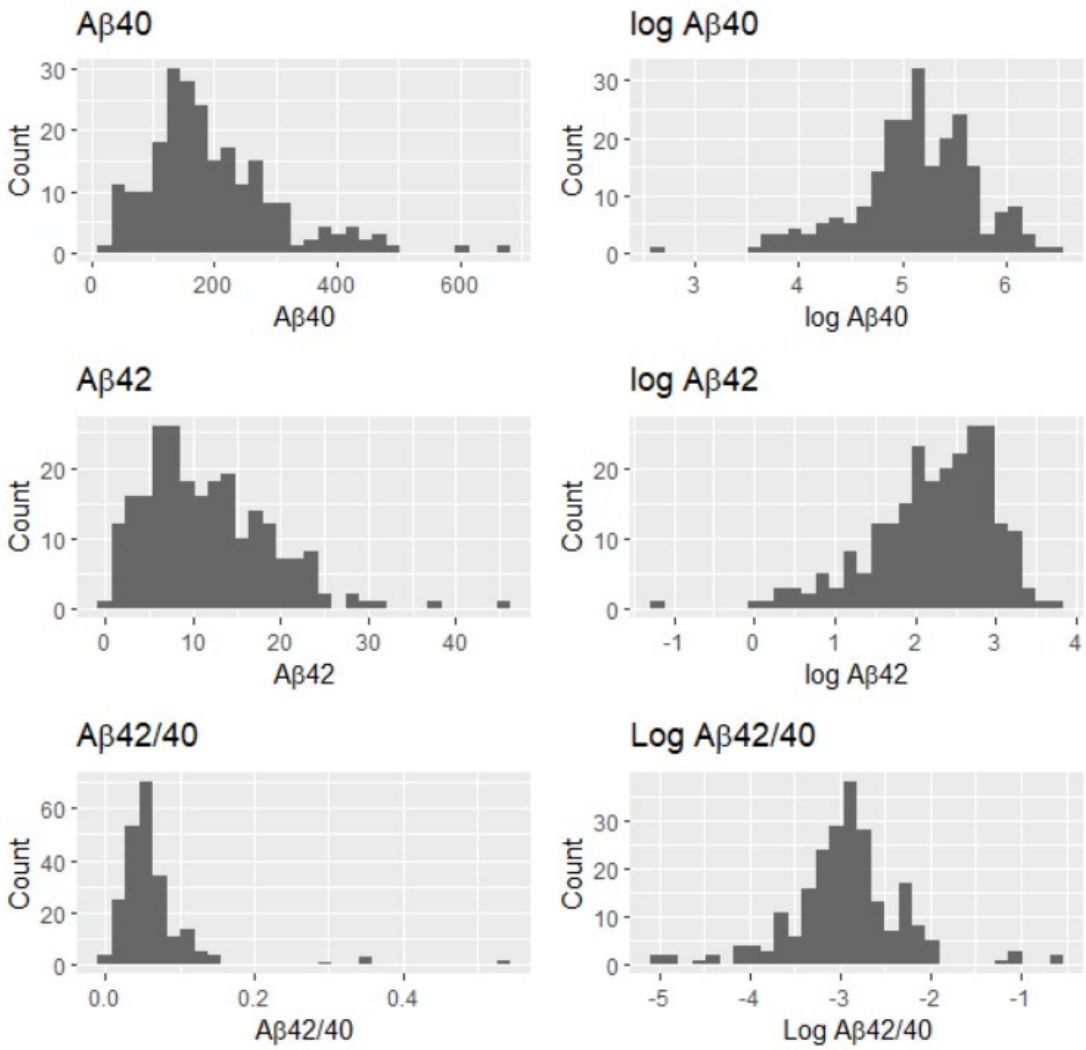
SD: Standard Deviation; IQR: Interquartile Range; Effective N: number of observations where biomarker is not missing out of possible 277 total individuals (2012/13 data). *MMP9 values are presented in thousands.

Table 2. Selected covariates by biomarker group

Covariates	Neurodegenerative	Vascular	Inflammatory
Age	X	X	X
APOE	X		
B12			X
Body mass index			X
Cancer			X
Cerebrovascular disease		X	X
COPD			X
Cardiovascular disease			X
Depression			X
Diabetes		X	X
Hypercholesterolemia		X	X
Hypertension	X	X	X
Gender	X	X	X
Smoker		X	X

Note: Sufficient adjustment sets representing selected covariates for each biomarker subgroup. APOE: Apolipoprotein E; COPD: chronic obstructive pulmonary disease.

Figure 1



Note: Histograms of biomarkers before and after log transformation. Remainder of biomarker histograms in Supplemental Materials. Aβ: amyloid beta.

Table 3. Included University of Kentucky Alzheimer's Disease Research Center baseline participant characteristics (N=237).

Variable	Summary
Age (mean ± sd)	82.69 ± 7.46
BMI (mean ± sd)	26.55 ± 4.61
Batch (2013/2018)	60 (22)
Gender (F)	171 (62)
APOE (any *e4* allele)	86 (31)
Cerebrovascular disease	23 (8)
Cardiovascular disease	58 (21)
Hypertension	167 (61)
Diabetes	39 (14)
Hypercholesterol	176 (64)
Smoker	126 (45)
B12 Deficiency	25 (9)
COPD	19 (7)
Cancer	66 (24)
Depression	52 (19)

Note: Unless otherwise stated in Variable column, statistics reported are number (%) and variables are coded as 0=never having condition and 1=ever having condition, unless otherwise stated.
BMI: Body mass index; APOE: Apolipoprotein E; COPD: Chronic obstructive pulmonary disease.

Table 4

Neurodegenerative/AD Biomarker Model Results*																		
Outcome:	log(Aβ40)			log(Aβ42)			log(Aβ42/Aβ40)			log(Tau)			log(Tau/Aβ42)			log(NfLight)		
Predictor	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val
APOE	0.042	0.084	0.615	-0.141	0.105	0.181	-0.194	0.079	0.015	-0.073	0.103	0.478	0.056	0.119	0.639	0.091	0.085	0.282
Age	0.013	0.006	0.033	-0.006	0.008	0.424	-0.019	0.006	0.001	-0.017	0.008	0.030	-0.011	0.009	0.221	0.039	0.006	<0.001
Hypertension	0.191	0.076	0.013	0.076	0.096	0.429	-0.136	0.072	0.059	0.202	0.095	0.034	0.113	0.109	0.300	0.057	0.077	0.464
Gender	0.051	0.077	0.510	0.058	0.097	0.552	0.018	0.072	0.800	-0.050	0.095	0.597	-0.107	0.109	0.331	0.129	0.078	0.099
Batch	-0.079	0.103	0.447	0.534	0.132	<0.001	0.671	0.097	<0.001	0.043	0.129	0.740	-0.509	0.149	0.001	0.662	0.106	<0.001
R-squared	0.075			0.117			0.363			0.046			0.063			0.189		
Effective N	227			233			225			235			233			235		

**NOTE: Model results for main analysis for Neuro/AD biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value.*

Table 5

Vascular Biomarker Model Results						
Outcome:	log(PlGF)			log(MMP9)		
Predictor	Est.	SE	P-val	Est.	SE	P-val
Age	0.014	0.016	0.371	0.030	0.012	0.015
CB	-0.545	0.429	0.205	-0.042	0.319	0.896
CV	0.249	0.245	0.312	-0.245	0.188	0.194
Diabetes	-0.248	0.283	0.381	0.054	0.218	0.805
Hypercholesterol	0.383	0.212	0.072	0.100	0.163	0.541
Hypertension	-0.304	0.213	0.155	0.028	0.162	0.864
Gender	0.091	0.200	0.648	0.036	0.153	0.815
Smoking	-0.275	0.192	0.154	0.133	0.147	0.365
Batch	0.525	0.271	0.054	1.371	0.211	<0.001
R-squared		0.063			0.174	
Effective N		226			232	

Note: Model results for vascular biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value.

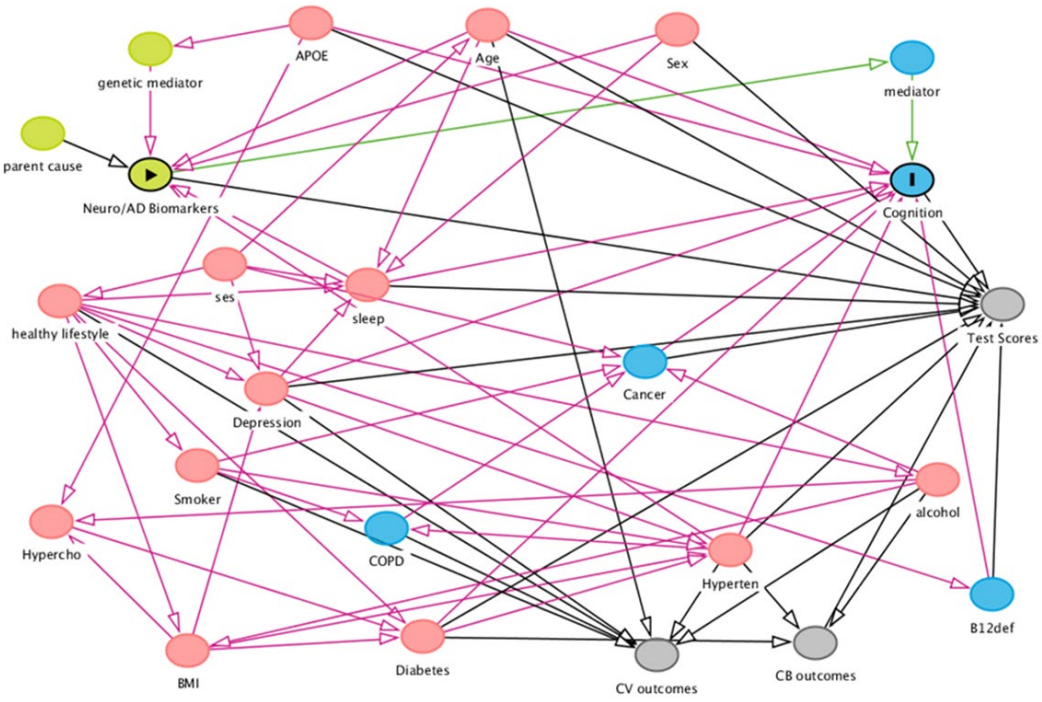
Table 6

Inflammatory Biomarker Model Results															
Outcome:	log(TNFα)			log(IL6)			log(IL8)			log(IL10)			log(IL1b)		
Predictor	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val
Age	0.002	0.009	0.845	0.015	0.010	0.131	0.017	0.011	0.137	0.021	0.008	0.006	-0.003	0.017	0.866
B12 Defficiency	0.017	0.183	0.924	0.327	0.206	0.115	-0.104	0.242	0.667	0.260	0.169	0.124	-0.617	0.380	0.106
BMI	0.003	0.011	0.782	0.020	0.013	0.128	-0.017	0.014	0.237	0.003	0.010	0.770	0.018	0.024	0.466
CB	-0.084	0.233	0.721	0.034	0.274	0.901	0.000	0.286	0.999	0.184	0.213	0.387	0.349	0.453	0.442
CV	0.199	0.132	0.133	0.195	0.148	0.190	-0.033	0.170	0.845	-0.002	0.115	0.986	-0.017	0.265	0.950
COPD	0.030	0.203	0.882	-0.260	0.229	0.258	0.163	0.251	0.516	-0.122	0.178	0.493	-0.203	0.412	0.622
Cancer	0.146	0.125	0.244	0.330	0.140	0.019	-0.090	0.158	0.569	0.196	0.108	0.071	0.332	0.255	0.195
Depresison	-0.164	0.135	0.226	-0.122	0.153	0.427	-0.046	0.168	0.785	0.159	0.119	0.181	-0.515	0.276	0.064
Diabetes	0.106	0.150	0.481	-0.068	0.169	0.687	0.000	0.191	0.998	0.069	0.131	0.599	0.023	0.298	0.938
Hypercholesterol	0.140	0.113	0.215	-0.093	0.127	0.463	-0.054	0.145	0.707	-0.071	0.099	0.472	0.007	0.228	0.974
Hypertension	0.012	0.113	0.919	0.080	0.127	0.529	0.029	0.143	0.840	-0.069	0.098	0.481	-0.369	0.228	0.107
Gender	0.051	0.106	0.628	-0.032	0.120	0.788	0.016	0.136	0.907	-0.133	0.093	0.157	-0.023	0.217	0.915
Smoking	-0.138	0.103	0.180	0.000	0.116	0.997	0.139	0.130	0.287	-0.043	0.090	0.630	-0.050	0.207	0.810
Batch	-0.294	0.146	0.046	0.314	0.166	0.059	0.625	0.181	0.001	0.110	0.128	0.389	-0.628	0.310	0.044
R-squared	0.075			0.082			0.088			0.1			0.09		
Effective N	225			226			204			225			201		

Note: Model results for inflammatory biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value.

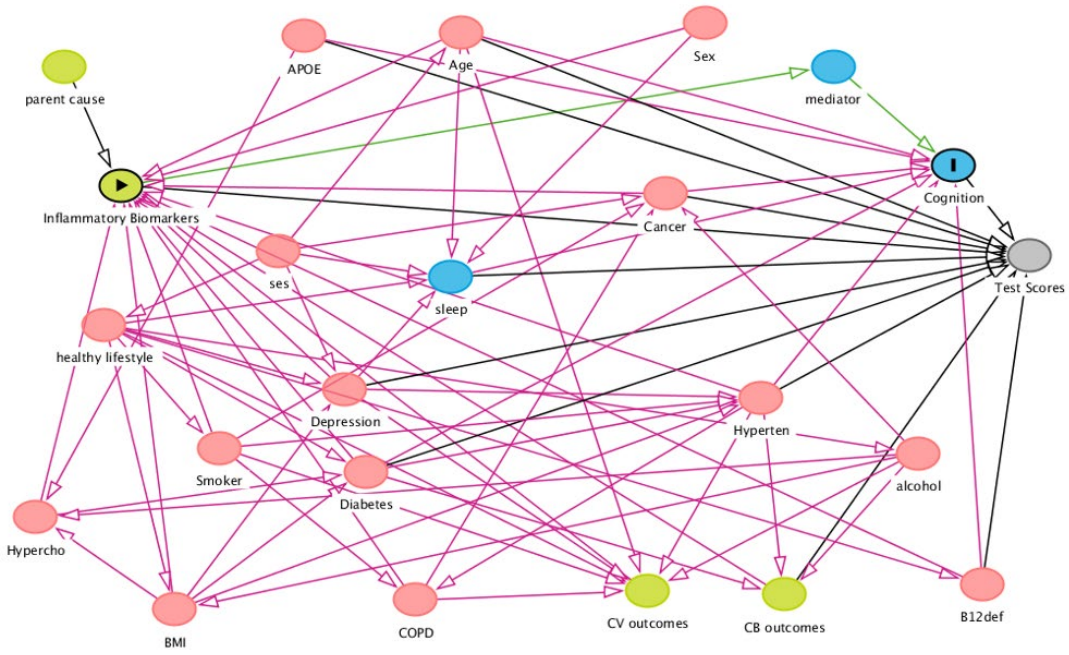
Supplemental Materials

Supplemental Figure 1



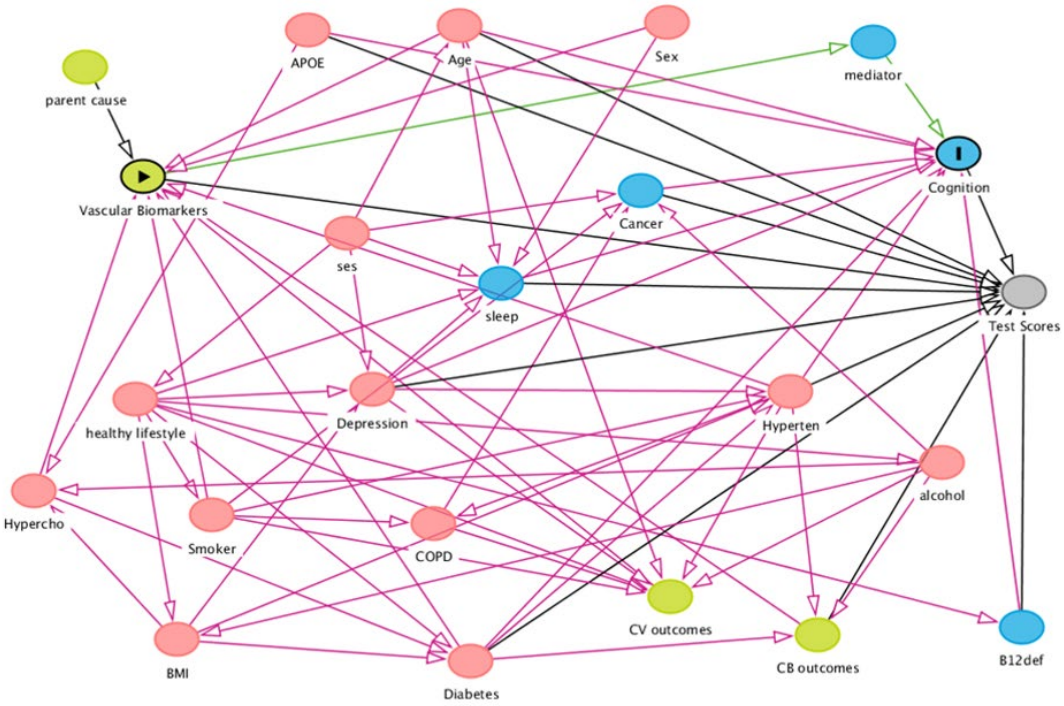
Note: Neuro/AD Biomarkers Directed Acyclic Graph (DAG) made with daggity.net. Neuro/AD Biomarker node is exposure, yellow nodes are ancestors of the exposure, Cognition node is the outcome, blue nodes are ancestors of the outcome, pink nodes are ancestors of exposure and outcome, gray nodes are other variables. Green arrow is a causal pathway, pink arrow is a biasing pathway, black arrows are connections between other variables.
Neuro/AD,: Neurodegenerative/Alzheimer's Disease; APOE: Apolipoprotein E; B12def: vitamin B12 defficiency; BMI: body mass index; CB outcomes: cerebrovascular outcomes; COPD: Chronic obstructive pulmonary disease; CV outcomes: cardiovascular outcomes; Hypercho: hypercholesterolemia; Hyperten: hypertension; ses: socioeconomic status.

Supplemental Figure 2



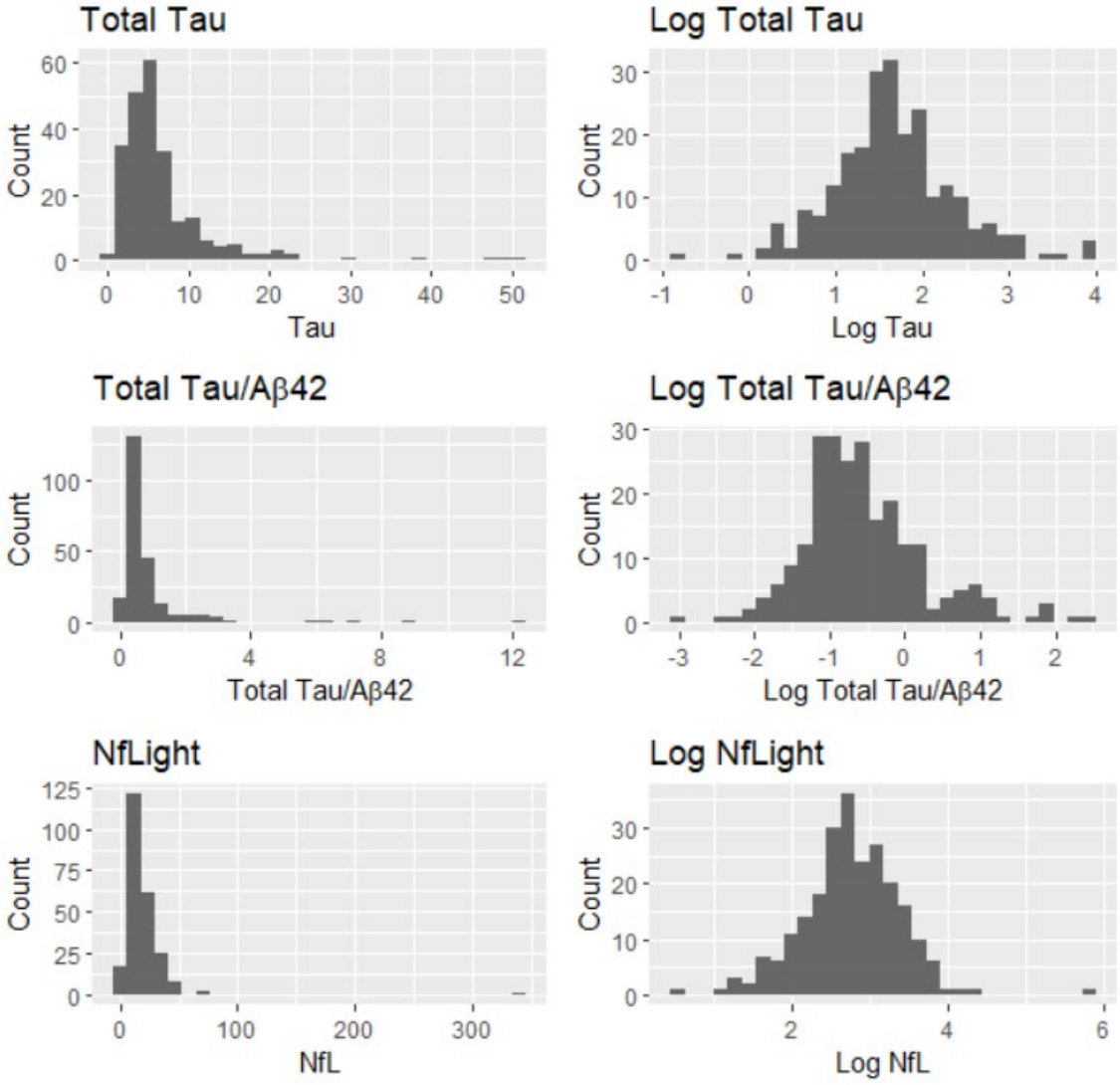
Note: Inflammatory Biomarker DAG.

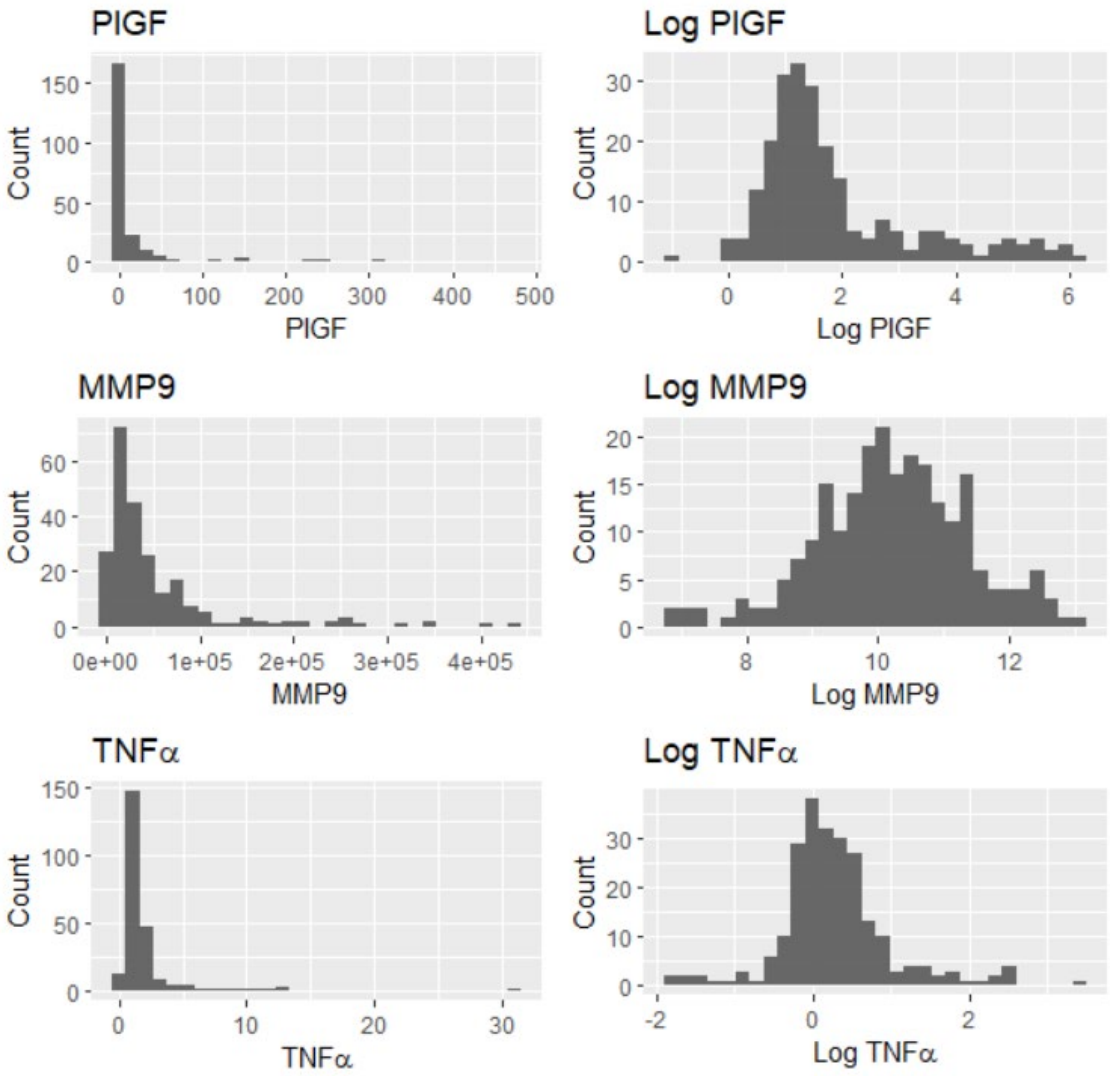
Supplemental Figure 3

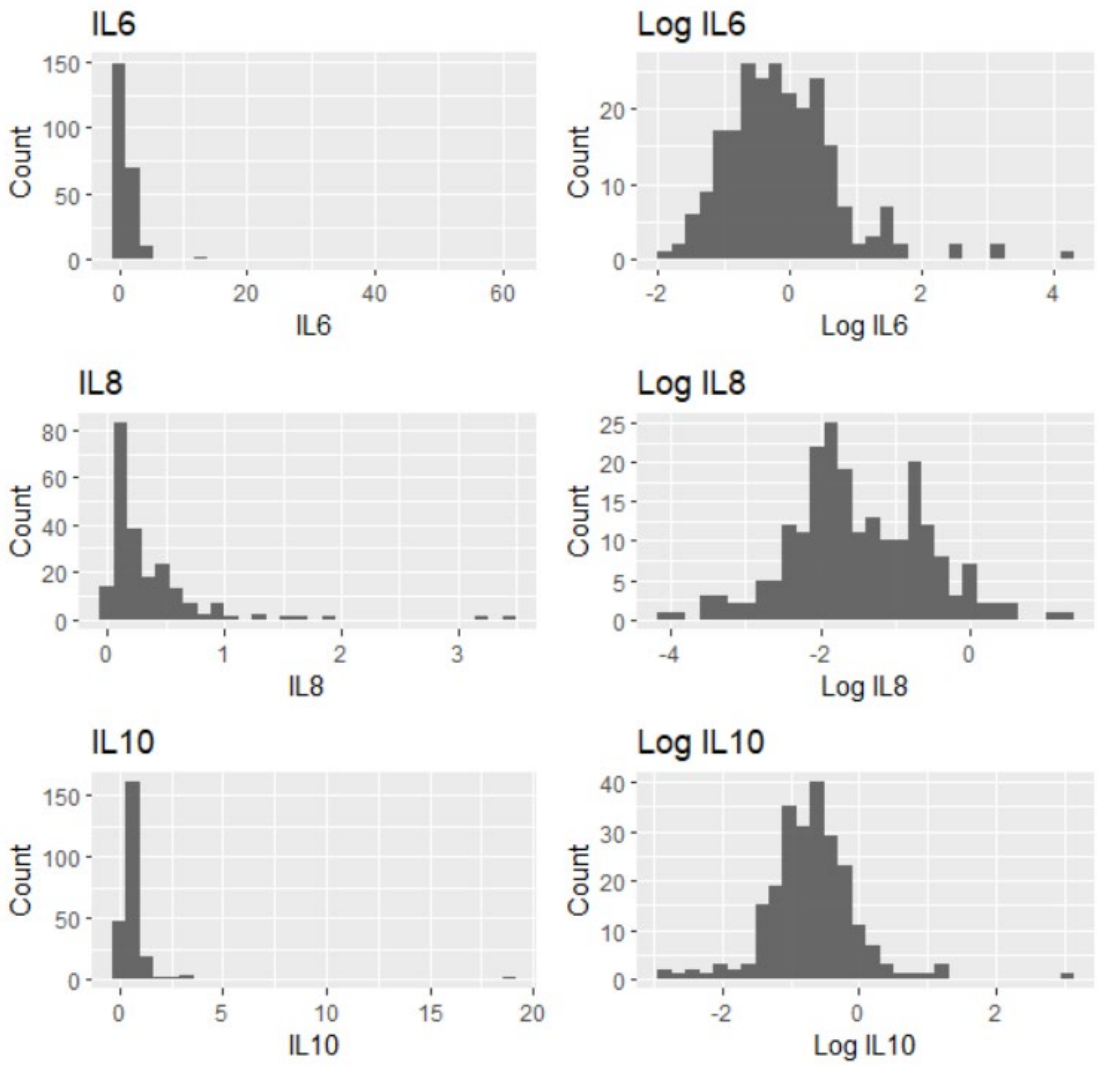


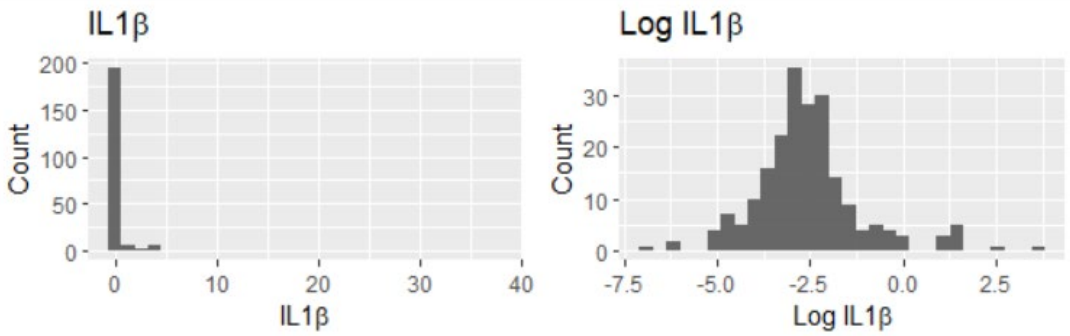
Note: Vascular Biomarker DAG.

Supplemental Figure 4




PIGF
Count
PIGF
Log PIGF
Log PIGF
MMP9
MMP9
Log MMP9
Log MMP9
TNFα
TNFα
Log TNFα
Log TNFα


IL6
Count
IL6
Log IL6
Log IL6
IL8
Count
IL8
Log IL8
Log IL8
IL10
Count
IL10
Log IL10
Log IL10



Note: Remaining biomarker histograms.

Supplemental Table 1

Biomarker Correlation Matrix (Visit 1)													
	Aβ40	**Aβ42**	**Aβ42/40**	**Tau**	**Tau/Aβ42**	**NfLight**	**PlGF**	**MMP9**	**IL6**	**IL8**	**IL10**	**IL1b**	**TNFα**
Aβ40	1	0.629	-0.278	0.271	-0.189	0.159	0.136	0.021	0.020	-0.025	0.096	-0.018	0.177
Aβ42	0.629	1	0.336	0.239	-0.350	0.308	0.115	0.124	0.011	0.053	0.052	-0.024	0.117
Aβ42/40	-0.278	0.336	1	-0.030	-0.183	0.125	-0.034	0.163	-0.012	0.095	-0.003	-0.020	-0.073
Tau	0.271	0.239	-0.030	1	0.482	0.067	0.072	0.044	-0.018	0.012	-0.018	-0.016	0.093
Tau/Aβ42	-0.189	-0.350	-0.183	0.482	1	-0.079	0.007	-0.008	-0.036	-0.016	-0.043	-0.018	-0.030
NfLight	0.159	0.308	0.125	0.067	-0.079	1	0.017	0.057	0.009	0.381	0.076	0.000	-0.017
PlGF	0.136	0.115	-0.034	0.072	0.007	0.017	1	0.082	0.276	-0.037	0.041	0.255	0.694
MMP9	0.021	0.124	0.163	0.044	-0.008	0.057	0.082	1	0.025	0.005	-0.050	0.278	0.070
IL6	0.020	0.011	-0.012	-0.018	-0.036	0.009	0.276	0.025	1	-0.002	0.135	0.149	0.179
IL8	-0.025	0.053	0.095	0.012	-0.016	0.381	-0.037	0.005	-0.002	1	0.093	-0.014	-0.031
IL10	0.096	0.052	-0.003	-0.018	-0.043	0.076	0.041	-0.050	0.135	0.093	1	-0.027	0.121
IL1b	-0.018	-0.024	-0.020	-0.016	-0.018	0.000	0.255	0.278	0.149	-0.014	-0.027	1	0.101
TNFα	0.177	0.117	-0.073	0.093	-0.030	-0.017	0.694	0.070	0.179	-0.031	0.121	0.101	1

Note: Correlation matrix for biomarkers using data from first visits (2012 or 2013).

Supplemental Table 2

Batch Effect Sizes	
Outcome	Effect Size
Aβ40	0.14
Aβ42	0.72
Aβ42/40	1.05
Tau	0.06
Tau/Aβ42	0.62
NfLight	1.05
PlGF	0.37
MMP9	1.17
TNFα	0.40
IL6	0.36
IL8	0.68
IL10	0.17
IL1b	0.43
TNFα/PlGF	0.68

Note: Effect sizes for batch coefficient for all main analysis models and post hoc analysis model (TNFα/PlGF). All outcomes are logged in models. Effect sizes presented are in absolute value.

Supplemental Table 3

Neurodegenerative/AD Biomarker Sensitivity Analysis Results																		
Outcome:	log(Aβ40)			log(Aβ42)			log(Aβ42/Aβ40)			log(Tau)			log(Tau/Aβ42)			log(NfLight)		
Predictor	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val
APOE	-0.018	0.085	0.830	-0.124	0.110	0.260	-0.116	0.072	0.108	0.162	0.128	0.206	0.286	0.114	0.013	0.010	0.101	0.918
Age	0.022	0.006	0.001	0.000	0.008	0.981	-0.022	0.005	<0.001	0.013	0.010	0.192	0.013	0.009	0.139	0.024	0.008	<0.001
Hypertension	0.045	0.076	0.552	0.082	0.099	0.407	0.019	0.064	0.766	0.049	0.115	0.670	-0.033	0.103	0.748	-0.021	0.091	0.817
Gender	0.015	0.078	0.850	0.112	0.102	0.274	0.068	0.065	0.300	0.027	0.118	0.822	-0.085	0.106	0.423	0.147	0.093	0.119
Batch	-0.066	0.095	0.484	0.655	0.124	<0.001	0.748	0.080	<0.001	0.628	0.144	<0.001	-0.027	0.129	0.836	0.359	0.114	<0.001
R-squared		0.089			0.154			0.452			0.1			0.045			0.082	
Effective N		183			189			183			189			189			189	

Note: Sensitivity analysis results for Neuro/AD biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value

Supplemental Table 4

Vascular Biomarker Sensitivity Analysis Results						
Outcome:	log(PlGF)			log(MMP9)		
Predictor	Est.	SE	P-val	Est.	SE	P-val
Age	0.022	0.018	0.224	0.047	0.014	0.001
CB	-0.566	0.366	0.124	-0.033	0.284	0.907
CV	0.394	0.266	0.139	0.111	0.204	0.587
Diabetes	0.248	0.315	0.433	-0.189	0.241	0.435
Hypercholesterol	0.086	0.222	0.699	0.139	0.170	0.412
Hypertension	-0.151	0.224	0.503	0.059	0.171	0.729
Gender	-0.117	0.227	0.607	-0.264	0.171	0.124
Smoking	-0.230	0.211	0.277	0.264	0.161	0.104
Batch	0.520	0.270	0.056	1.500	0.226	<0.001
R-squared		0.066			0.241	
Effective N		185			180	

Note: Sensitivity analysis results for Vascular biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value.

Supplemental Table 5

Inflammatory Biomarker Sensitivity Analysis Results															
Outcome:	log(TNFα)			log(IL6)			log(IL8)			log(IL10)			log(IL1β)		
Predictor	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val	Est.	SE	P-val
Age	0.004	0.012	0.725	0.015	0.013	0.253	0.017	0.014	0.233	0.008	0.012	0.487	0.040	0.023	0.089
B12 Defficiency	0.188	0.433	0.664	1.151	0.461	0.014	-1.082	0.454	0.018	1.376	0.404	0.001	1.021	0.749	0.175
BMI	-0.010	0.014	0.481	-0.012	0.017	0.492	-0.009	0.018	0.612	-0.025	0.015	0.092	0.020	0.029	0.495
CB	0.082	0.236	0.728	0.197	0.282	0.486	-0.113	0.289	0.695	0.180	0.239	0.453	-0.200	0.460	0.665
CV	0.174	0.164	0.289	0.175	0.193	0.367	-0.251	0.206	0.227	-0.102	0.169	0.547	-0.253	0.330	0.445
COPD	0.627	0.295	0.035	0.027	0.349	0.938	0.758	0.348	0.031	0.560	0.306	0.069	-0.318	0.573	0.580
Cancer	-0.001	0.163	0.996	0.338	0.190	0.077	-0.408	0.206	0.049	-0.078	0.165	0.637	0.768	0.314	0.016
Depresison	0.068	0.163	0.679	-0.057	0.188	0.761	-0.038	0.192	0.845	0.102	0.164	0.535	-0.351	0.327	0.284
Diabetes	0.087	0.198	0.659	0.119	0.231	0.608	-0.238	0.243	0.328	0.122	0.202	0.548	-0.021	0.405	0.959
Hypercholesterol	-0.002	0.139	0.989	-0.022	0.163	0.893	-0.054	0.179	0.765	0.034	0.143	0.814	-0.121	0.281	0.666
Hypertension	0.049	0.141	0.729	0.123	0.165	0.458	0.074	0.174	0.673	0.070	0.144	0.628	-0.415	0.284	0.146
Gender	-0.192	0.145	0.188	-0.108	0.168	0.523	0.175	0.176	0.321	-0.100	0.146	0.495	-0.135	0.287	0.638
Smoking	-0.108	0.134	0.420	-0.030	0.157	0.848	-0.196	0.164	0.233	-0.185	0.137	0.177	-0.150	0.270	0.581
Batch	0.170	0.175	0.334	0.158	0.204	0.439	0.657	0.201	0.001	-0.119	0.175	0.498	-0.248	0.369	0.502
R-squared		0.068			0.098			0.216			0.12			0.13	
Effective N		178			183			159			185			164	

Note: Sensitivity analysis results for Inflammatory biomarkers. Est: coefficient estimate; SE: Standard Error; P-val: p-value.

Supplemental Table 6

Post-Hoc Analysis			
Outcome:	log(TNFα/PlGF)		
Predictor	Est.	SE	P-val
Age	-0.010	0.014	0.472
B12 Defficiency	-0.310	0.291	0.288
BMI	-0.012	0.018	0.491
CB	0.052	0.374	0.890
CV	-0.103	0.207	0.618
COPD	0.360	0.322	0.264
Cancer	-0.094	0.198	0.635
Depression	-0.040	0.211	0.849
Diabetes	0.397	0.238	0.097
Hypercholesterol	-0.135	0.178	0.447
Hypertension	0.237	0.179	0.187
Gender	-0.031	0.168	0.854
Smoking	0.134	0.162	0.407
Batch	-0.781	0.232	0.001
R-squared	0.098		
Effective N	218		

Note: Model results for post hoc analysis of the ratio of TNFα and PlGF. Est: coefficient estimate; SE: Standard Error; P-val: p-value.

Supplemental Table 7

2012/13 Diagnosis	2017/18 Diagnosis			
	Normal	MCI	Demented	2012/13 Totals
Normal	190	38	9	237
MCI	-	12	21	33
Demented	-	-	7	7
2017/18 Totals	190	50	37	277

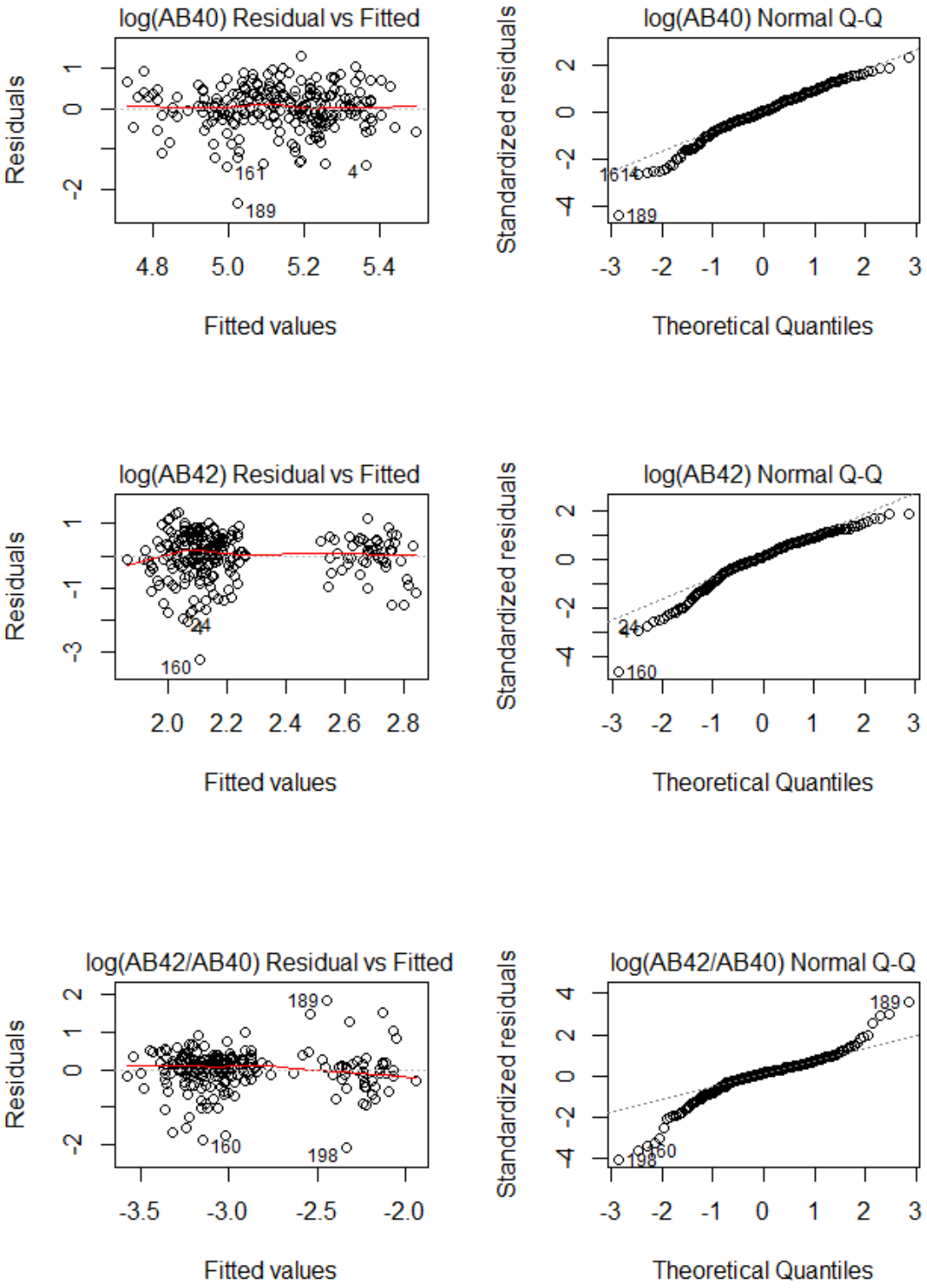
Note: Diagnosis transition table.

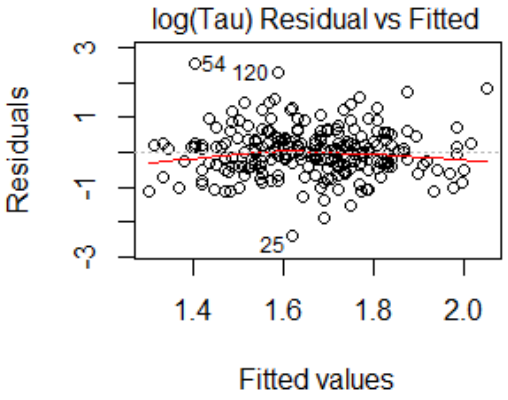
Supplemental Table 8

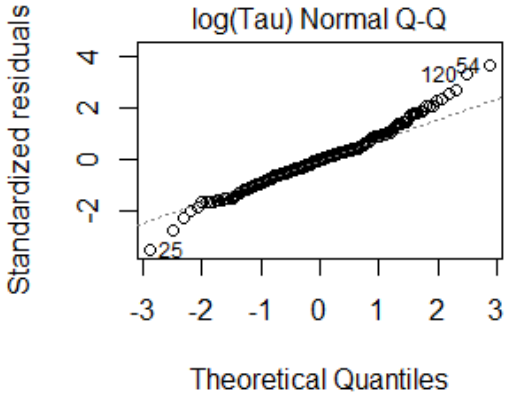
	Batch	
Demographics	0	1
Gender (F)	0.59	0.7
Apoe (+)	0.31	0.32
Smoker (ever)	0.44	0.47
Age (yrs)	84.7(6.2)	78(7.6)
Education (yrs)	16.7(2.7)	16.4(2.8)

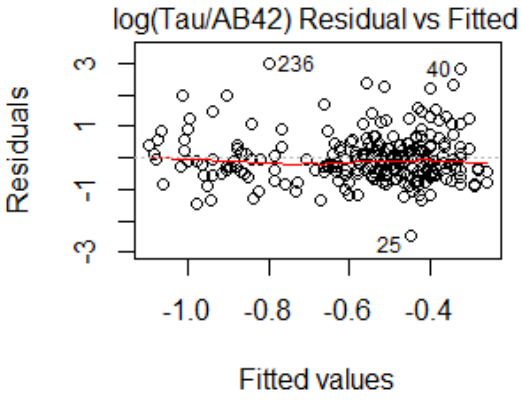
Note: Batch demographics table. Batch 0 = 2012-2017, Batch 1 = 2013-2018.
Age and education present mean(standard deviation), otherwise values are proportions.
APOE: Aplipoprotien E.

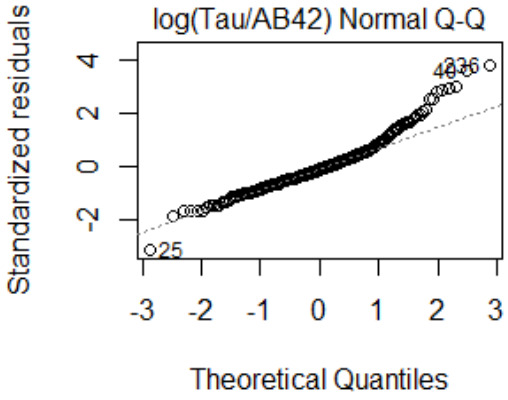
Supplemental Figure 5

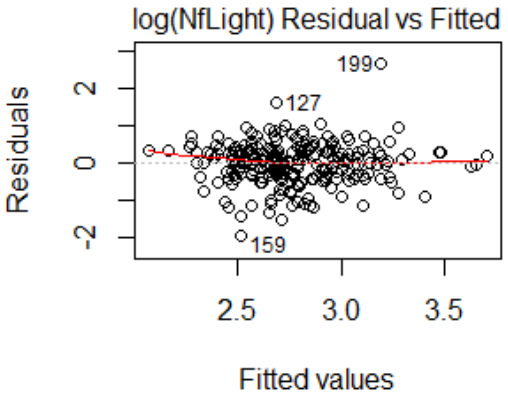


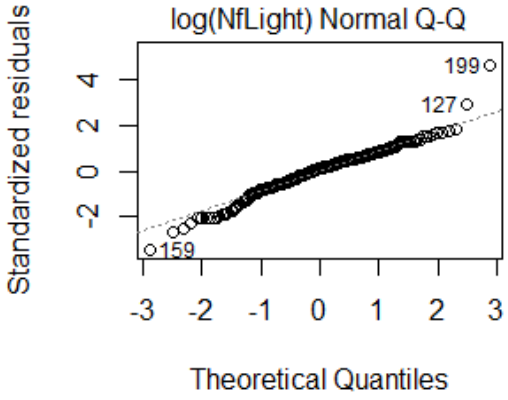

log(Tau) Residual vs Fitted
Residuals
Fitted values
54
120
25
1.4
1.6
1.8
2.0
-3
-1
1
3

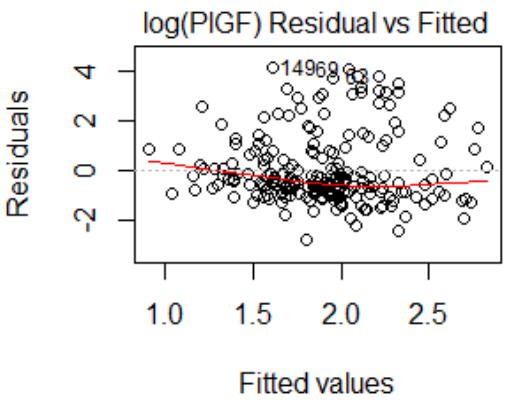

log(Tau) Normal Q-Q
Standardized residuals
Theoretical Quantiles
120
54
25
-3
-2
-1
0
1
2
3
4

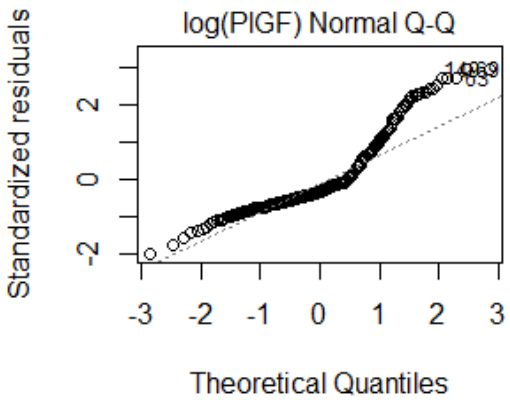

log(Tau/AB42) Residual vs Fitted
Residuals
Fitted values
236
40
25
-1.0
-0.8
-0.6
-0.4
-3
-1
1
3

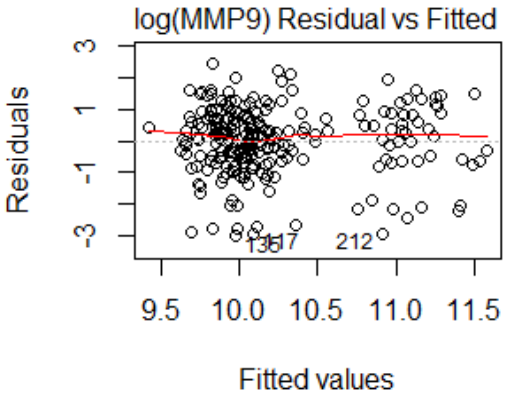

log(Tau/AB42) Normal Q-Q
Standardized residuals
Theoretical Quantiles
25
-3
-2
-1
0
1
2
3
4

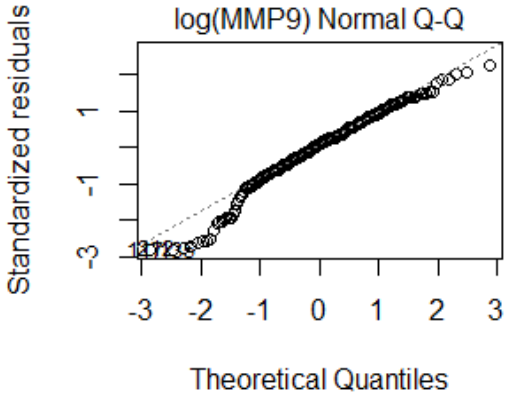

log(NfLight) Residual vs Fitted
Residuals
Fitted values
199
127
159
2.5
3.0
3.5
-2
0
2

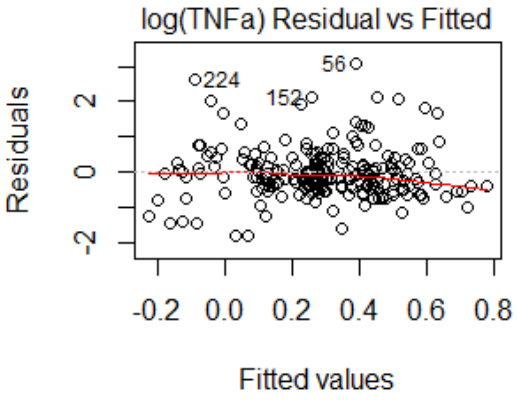

log(NfLight) Normal Q-Q
Standardized residuals
Theoretical Quantiles
199
127
159
-3
-2
-1
0
1
2
3
4

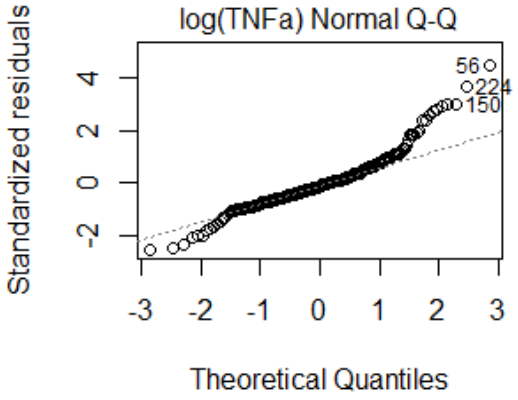

log(PlGF) Residual vs Fitted
Residuals
Fitted values
1.0
1.5
2.0
2.5
-2
0
2
4

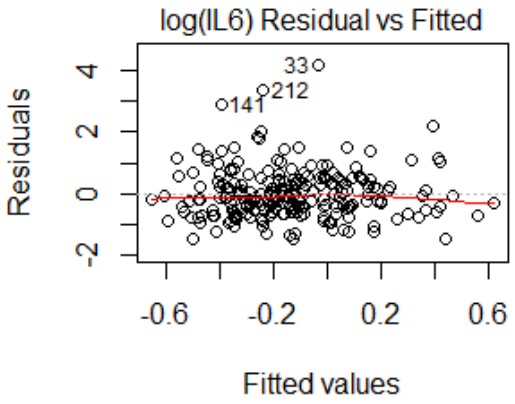

log(PlGF) Normal Q-Q
Standardized residuals
Theoretical Quantiles
-3
-2
-1
0
1
2
3

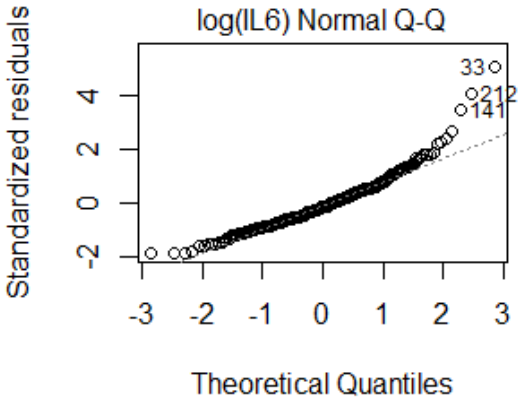

log(MMP9) Residual vs Fitted
Residuals
Fitted values
9.5
10.0
10.5
11.0
11.5
-3
-1
1
3

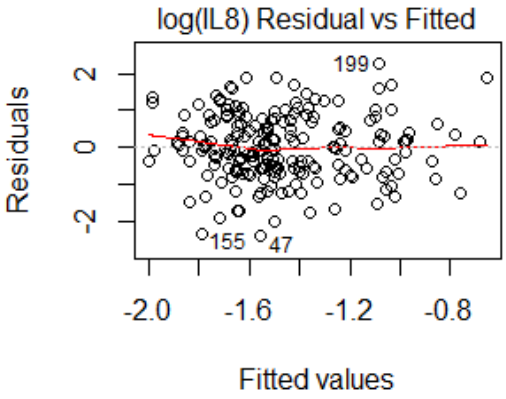

log(MMP9) Normal Q-Q
Standardized residuals
Theoretical Quantiles
-3
-2
-1
0
1
2
3

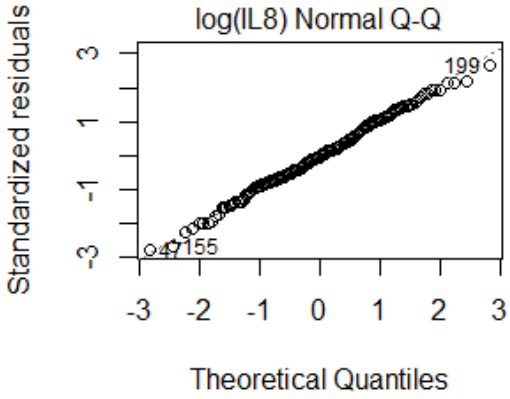

log(TNFa) Residual vs Fitted
Residuals
Fitted values
-0.2
0.0
0.2
0.4
0.6
0.8
-2
0
2

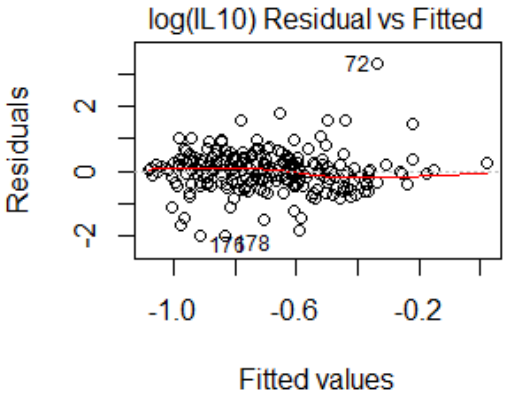

log(TNFa) Normal Q-Q
Standardized residuals
Theoretical Quantiles
-3
-2
-1
0
1
2
3
-2
0
2
4

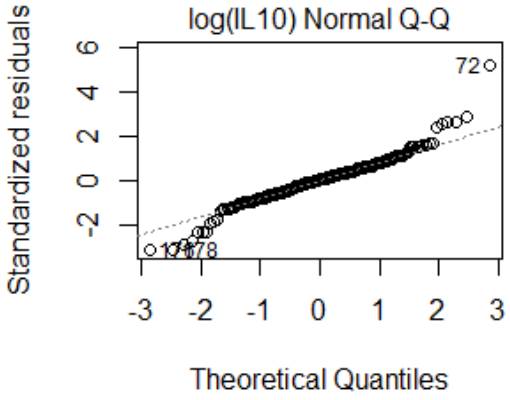

log(IL6) Residual vs Fitted
Residuals
Fitted values
33
212
141

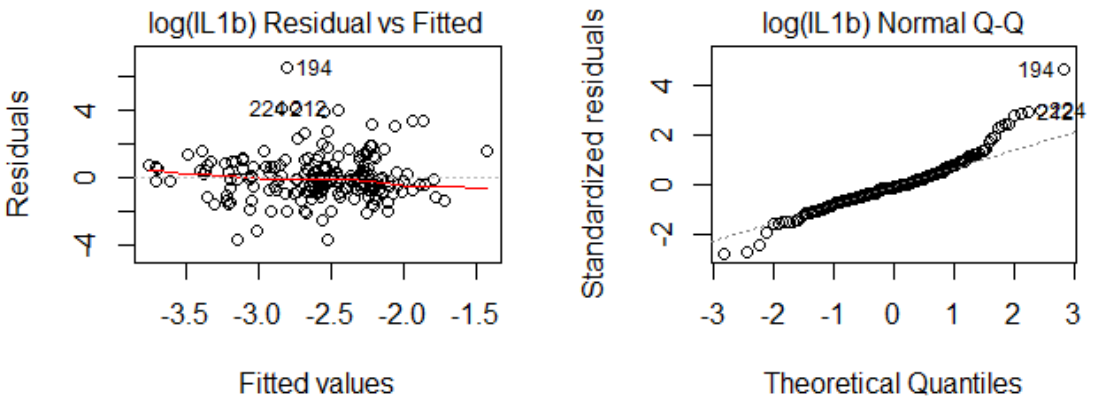

log(IL6) Normal Q-Q
Standardized residuals
Theoretical Quantiles
33
212
141


log(IL8) Residual vs Fitted
Residuals
Fitted values
199
155
47


log(IL8) Normal Q-Q
Standardized residuals
Theoretical Quantiles
199
47
155


log(IL10) Residual vs Fitted
Residuals
Fitted values
72


log(IL10) Normal Q-Q
Standardized residuals
Theoretical Quantiles
72



Note: Model assessment plots for main analysis.

CHAPTER 3

Introduction

Cognitive decline due to Alzheimer's disease (AD) is characterized by memory loss, reduced motor function, inability to think clearly, decrease in processing speed, and more [95]; these changes generally occur over decades. Although two anti-amyloid, disease modifying therapies have been recently approved by the US Food and Drug Administration (Aduhelm [96-97] and Leqembi [98-99]), there is much controversy about the clinical effectiveness of these treatments. Prevention remains the most promising strategy for addressing the immense public health burden of dementia, which can be characterized by the estimated $355 billion spent on health and long-term care of dementia patients and an estimated $256.7 billion worth of care provided by family members and other unpaid caregivers in 2021 and 2020, respectively [100], not to mention the emotional, mental, and physical burden that effects family members daily. There is much work being done in relation to identification and preventative measures, one avenue of which is the development of prognostic biomarkers.

Multiple candidate biomarkers have been identified for use in studies of cognitive decline. Neuroimaging and cerebrospinal fluid (CSF) biomarkers have shown some promising results for diagnostic biomarkers, but their widespread use is hindered by their cost and invasive nature [64-65,67,101-102]. The desire for a cost-effective, easily accessible, and minimally invasive biomarker has led researchers to focus on blood plasma biomarkers [67,101-103]. Given the complexity and heterogeneity of the neuropathology that underlies dementia [104], it is likely that multiple biomarkers will be necessary to best detect and predict disease progression. While the typical study of plasma biomarkers is likely to examine 4-5 separate biomarkers [87-88,90-92,103,105-106], studies often examine fewer [107-110], and rarely examine more [111].

Many studies in this realm use cognitive diagnosis as their outcome [87-88,90,106], while some use various cognitive test scores, mainly the Mini Mental State Exam (MMSE) [87,105].

Studies have used a wide variety of analytical methods, with varying levels of complexity, to examine biomarkers in relation to cognition and related demographics and health factors. Popular forms of analysis include receiver operating characteristic (ROC) area under the curve (AUC) [87-88,105,108,110]; various types of linear, logistic, and survival regression models [87-88,90,105-106,108]; as well as meta-analysis and review papers [101,103,107,109,111-113]. However, available literature using more sophisticated statistical methods in ADRD research is limited and application to plasma biomarkers even more so.

Plasma biomarkers have been found to be significantly associated with AD and other measures of cognitive decline in a myriad of ways. Plasma measures of Aβ40 [114], Aβ42 [106], Aβ42/40 [87-88,114], total tau [90,87,114], p-tau [87-88], NfLight [87-88,106,114], and IL6 [90] have all shown significant association with dementia diagnosis, to name just a few. MMP9 has been associated with blood-brain barrier breakdown, which contributes to cognitive decline [107]. In longitudinal data, plasma NfLight has been significantly associated with MMSE measures over time, and "more abnormal plasma measures were associated with an accelerated decline in MMSE scores," [87]. In addition to their relationship with cognitive decline, many biomarkers have been significantly associated with each other, supporting our idea that they may work in tandem to influence cognition. Aβ42 has been statistically significantly linked to total tau, TNFα, and IL6 [90], p-tau has shown association with Aβ measurements [87], and significant associations have been found between the inflammatory markers IL6, IL10, and TNFa [90].

Thus, the aim of the current study is to evaluate the usefulness of a panel of Quanterix HD-X SiMOA plasma biomarkers, measured in a cohort of older adult research volunteers enrolled at the University of Kentucky Alzheimer's Disease Research Center (UKADRC), to predict current and future cognitive performance. Participant data were obtained from two visits, 5 years apart. The longitudinal aspect of our study data gives us the opportunity to evaluate the potential of linear combinations of the biomarkers as predictors for three well-known and widely used cognitive exams: Mini Mental State Exam (MMSE), California Verbal Learning Test Long Delay Recall

(CVLT-LD), and Trail Making Test (Trails). Each of these instruments target different cognitive domains: global, memory, and executive function.

Methods

Setting

Data for this study was drawn from the UKADRC community based longitudinal cohort, which evaluates participant cognition and overall health approximately annually. This subset of participant data was selected as described by Estepp et. al. [115]. Briefly, included participants with cognitive assessment who completed UKADRC study visits in either 2013/2018 or 2014/2019 and had available plasma biomarker data. Both sets of years were combined, and 2013/2018 data was used if a participant was present in both sets, retaining two observations per participant. No exclusions were made based on cognitive status or medical conditions. The University of Kentucky Institutional Review Board approved all UKADRC study procedures, and all participants gave written informed consent.

Plasma Biomarkers

The UKADRC Biomarker Core provided data from a panel of 12 blood plasma biomarkers, comprising the following biomarker subtypes: neurodegenerative, vascular, and inflammatory. This study, which was performed following the conclusion of [115], includes the same battery of biomarkers, with the addition of phosphorylated-tau(p-tau)181 to the neurodegenerative subgroup. A listing of the biomarkers is provided in [Table 1]. Detailed methods for these biomarker assays have been published previously Estepp et al [115]. Briefly, banked participant EDTA plasma samples were run by study year, with the samples from 2013/2018 and then 2014/2019 being run at the same time to minimize batch effects. Commercially available Quanterix SiMOA kits were used to measure all biomarkers on the HDX analyzer. Multiplex kits were used to measure amyloid

beta 1-40 (Aβ40), amyloid beta 1-42 (Aβ42), and total tau. Uniplex kits were used to measure the rest of the biomarkers. We previously found no strong associations between the biomarkers and various medical conditions or demographic characteristics with linear models, leading to minimal worry of confounding by these factors in the current analysis [115].

Cognitive Measures

UKADRC participants are administered the National Alzheimer's Coordinating Center's Uniform Data Set (UDS) Neuropsychological Battery approximately annually, in addition to the CVLT. However, the UDS Neuropsychological Battery was substantially revised in March 2015 with UDS 3.0. Thus, only some measures were available at both time points in the current study.

Our first selected outcome, MMSE [116], has been widely used in the field of AD research to measure global cognitive function. MMSE was removed from the UDS in 2015 when v3.0 was implemented [117], but the UKADRC has continued to collect this measure. MMSE consists of 16 questions, comprising oral, written, and action elements, totaling 30 possible points [116]. The length and breadth of this exam contributes to its reputation as a valid measure of global cognition [118]. Typical scoring of the MMSE subtracts the number incorrect responses, or points lost, from the 30 total points, effectively counting correct responses. For the current study, we chose to use the number of incorrect responses as the outcome. This facilitated modeling using a Poisson count distribution (see Statistical Analysis section below), without quantitatively changing its meaning (since the MMSE has a strict range of values and specified scoring system, examining the number of correct answers is the same as examining the number of incorrect answers).

The CVLT assesses the episodic memory domain [119]. It is thought to be sensitive to early, more subtle impairment, unlike the MMSE [120], which was developed to detect more serious impairments. The CVLT exam has two lists of words, A and B, both 16 words long, with 4 words from 4 different categories presented in a pseudo-random order [121]. List A is presented to the individual 5 times, followed by list B immediately after the fifth presentation of list A. Various

CVLT scores are extracted based on the patient's ability to recall the lists immediately, multiple times, and after a period of time (long delay recall). For this study, we selected the long delay recall (CVLT-LD) score, which captures the number of correct words recalled from list A after 20 minutes. Raw CVLT scores are converted to standard scores based on participant age and sex; CVLT-LD is operationalized as a z-score based on a reference population of cognitively intact adults [122].

To capture executive function, our third outcome comes from the Trail Making Test (TMT). The TMT has two parts, A and B. TMT A asks the participant to connect a series of dots numbered 1-25 in ascending order and the time elapsed in seconds is recorded as the score; maximum allowed time to complete is 180 seconds. TMT B asks the participant to connect a series of dots, numbered 1-12 and lettered A-L (24 dots total), in ascending order, alternating numbers and letters. Time elapsed in seconds is again recorded as the score; maximum time allowed to complete is 300 seconds [123-127]. TMT A and B are often used separately [123-128], but in the current study we chose to use TMT B-A as a measure for the extra difficulty in TMT B while controlling for TMT A (processing speed) for each individual [127].

Finally, in a post-hoc analysis, clinical cognitive diagnosis was used as an outcome. UKADRC clinicians followed explicit guidelines for diagnoses of Normal, Mild Cognitive Impairment (MCI), and Dementia [70].

Study Design

To assess the diagnostic and prognostic utility of the biomarkers, we constructed four iterations of the data. We speculated that in addition to using baseline levels to predict future cognitive status, the change in biomarkers over the 5-year study period may explain cognition, as well as their base levels. To this end, we constructed two cross-sectional studies and two longitudinal studies. In the first cross-sectional study, we included only information from Study Year 1 (2013 or 2014; dataset 1); the second cross-sectional included only Study Year 5 information

(2018 or 2019; dataset 2). The first longitudinal study used Study Year 1 biomarker and covariate data, and outcome data from Study Year 5 (dataset 3). The second longitudinal study used Study Year 5 covariates and outcomes but incorporated biomarker data from both Study Year 1 and Year 5 (dataset 4).

Evaluating the relations among biomarkers and outcomes in these four study designs provides results applicable in a variety of possible clinical settings. Results from the cross-sectional studies (datasets 1 and 2) provide information on associations of outcomes with variables, and dataset 2 results can be thought of as a sensitivity analysis for the results from dataset 1. Results from the first longitudinal study (dataset 3) may be the most relevant to clinicians, who often wish to use a participant's current state to predict their future outcomes. Finally, results from the second longitudinal study (dataset 4) may provide more insight toward disease progression if the change in biomarkers proves to be influential in explaining cognitive changes.

Statistical Analysis

The main interest for this analysis is to assess the relationship between the biomarkers and cognitive tests. We hypothesized that information from biomarkers needs to be considered collectively, as opposed to independently. In order to produce a relatively simple model that was not affected by collinearity but still included the most information from the biomarkers, we sought an appropriate dimension reduction technique.

We used partial least squares (PLS) analysis for dimension reduction because PLS uses the outcome in the formation of its components, unlike principal components analysis, deriving linear combinations of covariates that explain the variation in the outcome [129]. PLS thus achieves the desired dimension reduction while producing combinations of covariates that are related to the outcome.

In addition to PLS for dimension reduction of the biomarkers, we used Lasso to select covariates. The initial set of variables included in the Lasso was based on the confounders identified by previously developed Directed Acyclic Graphs [115]. Age (in years), gender, education (in years), Apolipoprotein E (*APOE)*, any e4 alleles vs none, and batch (13/18 vs 14/19), as described in [115]), were forced to stay in each of the Lasso models as critical prognostic factors for cognitive outcomes. In the Lasso procedure the number of variables included in each model (in addition to the 5 required variables) was determined via cross validation selection of the lambda parameter (default 10-fold validation for the R package was used) [14,130]. The lambda parameter in the Lasso procedure is a weight for the penalty factor applied to the covariates that measures the amount of simplification that occurs in the model through variable removal [14]. The "best" lambda value produces a model (using certain covariates) that has the smallest mean squared error (MSE). The "second-best" lambda value produces the model with the smallest number of covariates while maintaining a MSE within one standard error of the smallest MSE. To determine which lambda value to choose, which determines how many variables are to be included in the model, we evaluated plots of MSE by lambda values. The "second-best" lambda value was chosen for each model. This value produced the simplest models that performed reasonably well (via MSE), as the plots showed that the tradeoff of using lambda that improved MSE to its smallest value was never a drastic change in MSE, therefore it did not justify adding sometimes numerous variables to the models [Supplemental Figures 1-3].

After identifying covariates with the Lasso procedure, the parameters of interest from each of the four study designs were approximated with linear models including only the variables identified with the Lasso. Residuals from each of these models were taken and used as the data for the PLS procedure to ensure that the biomarker linear combinations would consider any other variables that would be in the final model and contribute information toward the outcome. We used linear models to produce the residuals for the PLS procedure for two reasons. First, we wanted to include any variables identified by the Lasso procedure in their full capacity, whereas the resulting

Lasso models may have penalties placed on some variables. Second, and perhaps more pertinent, the Lasso procedure has a reduced effective sample size because it requires observations to be complete cases when using all possible variables. By running our linear models, using complete cases only for the variables identified as useful by the Lasso, we can achieve a slightly larger sample size by not removing observations that are incomplete only for variables that were not included in the variable set. Once the PLS components for each model were identified, a final model for each study design-outcome combination was fit to the data including appropriate variables identified in the Lasso procedure and of appropriate model type (linear or Poisson). The statistical analysis process is outlined in [Figure 1].

Four data sets and three primary outcomes produced 12 sets of model results to evaluate. Participant age, gender, education, batch, and *APOE* were included in each model as covariates, in addition to identified PLS components.

Post Hoc Analyses

Two post hoc analyses were completed. First, a forward selection procedure was done for each outcome-dataset combination (starting with the 5 variables required to be in each model) to see if this simple procedure would add similar or differing variables to the models as compared to the Lasso-PLS process from the main analysis.

The second post hoc analysis followed the procedures outlined for the main analysis and first post hoc analysis but used clinical cognitive diagnosis (collapsed to normal vs. MCI or demented) as the outcome. Although the outcome was binary, a linear model was still used to produce the residuals for the PLS stage of the analysis, which is defensible given posterior classification above or below 0.5 [131]. This analysis was done to assess whether results for cognition aligned with the cognitive proxy outcomes used in the main analysis. The final models were changed to logistic regression to reflect the nature of the outcome.

Software

A 5% significance level was used when interpreting results. Analyses were performed with R version 4.1.2 in RStudio, using packages glmnet, tidyverse, dplyr, r2symbols, ggplot2, gridExtra, readxl, and haven [86,130,132-137].

Results

The analytic sample included 277 participants. Our sample was 64% female, 94% white, with an average age of 76.29 years at Study Year 1, and an average educational attainment of 16.67 years. Approximately 34% had at least one *APOE* e4 allele [Table 2]. About 20% of the individuals who were clinically Normal at Study Year 1 progressed to either MCI or Dementia at their Study Year 5 follow-up visit. Approximately 70% contributed biomarker information from their 2013/2018 visits, with the remaining 30% from the 2014/2019 visits. Results of the modeling procedures are given below.

Step 1: Lasso Procedure

The Lasso procedure identified few variables to include in each of the models beyond the pre-specified covariates (age, gender, education, *APOE*, and batch) [Table 3]. No additional covariates were selected for CVLD-LD; depression was selected for all four MMSE models; and the Trail B-A models selected no variables for dataset 1, diabetes for dataset 2, and hypertension for datasets 3 and 4.

Step 2: Linear Modeling Procedure

All cognitive test outcomes were approximated using linear models including the pre-specified covariates and any Lasso identified variables. Residuals from each model were extracted for the PLS procedure.

Step 3: Partial Least Squares Procedure

PLS analyses were run on the residuals from the previous step for each outcome-dataset combination. Validation plots showing mean squared error of prediction (MSEP) via cross validation, per number of components, showed that the addition of even one component for all models increased model error without improving accuracy [Supplemental Figures 4-6]. Biomarker coefficients for the first component in each model are presented in Table 4.

Step 4: Final Models

Final models for each outcome-dataset combination were run with the covariates as described above, and the first components identified by PLS. Linear models were used for CVLT-LD and Trails B-A, while MMSE errors used a Poisson generalized linear model. None of the biomarker component variables reached statistical significance [Table 5]. Batch did not reach significance for any of the outcome-dataset combinations. Among the other four prognostic variables purposely included in each model, age reached significance for CVLT-LD dataset 1, MMSE dataset 1, and Trails B-A datasets 2, 3, and 4; gender reached significance for CVLT-LD dataset 3, and MMSE datasets 1, 3, and 4; education reached significance for CVLT-LD datasets 1 and 3, and MMSE datasets 1, 2, 3, and 4; and *APOE* reached significance for MMSE datasets 1, 2, 3, and 4, and Trails B-A dataset 3. Depression, included in the MMSE models via the Lasso procedure, reached significance for all 4 datasets. For the Trails B-A models, neither diabetes (included for dataset 2) nor hypertension (included for datasets 3 and 4) reached significance.

Post Hoc Analyses

Forward selection was performed on each outcome-dataset combination, starting with the covariates as described above, to compare results with PLS findings. All covariates and biomarkers were available for selection (but not the component variable identified with PLS). This procedure,

based on model AIC, did not add any available covariates or individual biomarkers to the starting variables for any of the 12 models.

Following the completion of the main analysis and forward selection assessment, a post hoc analysis was done using the same 4 datasets for the outcome of clinical cognitive diagnosis. Cognition was specified as Normal vs. any cognitive decline due to small cell sizes for non-Normal diagnoses (Impaired, MCI, and Dementia). The same process followed in the main analysis [Figure 1] was carried out for this outcome.

The Lasso procedure, where the "second-best" lambda was again used for variable selection [Supplemental Figure 7], added no variables beyond the required five for dataset 1, and depression was selected for datasets 2, 3, and 4. Linear regression models were run to acquire residuals for the PLS procedure. PLS analysis was run for each dataset, and the first component was pulled for the next analysis [Supplemental Table 1]. Logistic regression models were run for each of the 4 datasets, including the pre-specified variables, Lasso identified variables, and the first component from the PLS procedure. The first PLS component variable did not reach significance for any of the datasets [Table 6]. Among the prognostic variables included in each model, batch and education did not reach significance for any of the 4 datasets, while age reached significance for datasets 2 and 3, gender reached significance for datasets 2, 3, and 4, and *APOE* reached significance for dataset 2. Depression, included in only datasets 2, 3, and 4 via the Lasso procedure, reached significance for all 3. This process was again followed by a standard forward selection procedure, which again did not add any variables to the models.

Discussion

We employed PLS analysis, in conjunction with variable selection via the Lasso procedure, with the goal of identifying informative linear combinations of plasma biomarkers in relation to our cognitive outcomes. While PLS analysis has been used in AD research, it is often used in studies

with neuroimaging outcomes [138-141] or without the use of potential AD biomarkers [138-139,142]. Similar literature that does utilize biomarkers with the PLS method typically uses very few or focuses on CSF biomarkers [128,140-141]. Our study makes a novel contribution in its use of many plasma biomarkers as predictors of non-neuroimaging cognitive outcomes.

We evaluated the possible use of linear combinations of biomarkers for the prediction of three different cognitive outcomes, considering some important patient characteristics. Age, gender, education, and *APOE* were included in each model because they have been prognostic in dementia research [74,143-145]. Batch was included in each model, even though all measures were taken on the same machines at the same facility, because our previous research showed that batch can still affect biomarker values [115], though it proved to be insignificant in all models in this study. Linear combinations of the biomarkers studied here, identified through PLS, were not shown to be useful in predicting cognitive outcomes in any setting we studied (for none of the cognitive tests or outcomes, neither cross-sectionally nor longitudinally). Following the Lasso-PLS procedure, forward selection was done as a more standard analysis for comparison, however, this also did not produce any significant associations.

As previously stated, many plasma biomarkers have been found to be statistically significantly associated with cognition, as well as each other [87-88,90,106-107,114]. While it was our hope to make clearer the relationships these biomarkers have with cognition and each other through the identification of influential linear combinations of biomarkers, we were unable to identify any statistically significant combinations. The reason for the lack of significant findings in this study cannot be easily determined. The sophisticated methodology we followed for this study is theoretically defensible but perhaps too convoluted for the relationships the biomarkers hold with cognitive outcomes. Additionally, a larger, more diverse sample may facilitate the discovery of more clinically significant relationships.

Despite the negative results, these findings are valuable and important to share. Research publications typically produce at least some positive findings, while null results are much less

common [146]. The presentation of only positive results for publication can lead to loss of time and money due to repetition of ineffective studies because of publication bias [147]. Our desire is to make advancements in the study of dementia, whether through our own findings or through guiding future research.

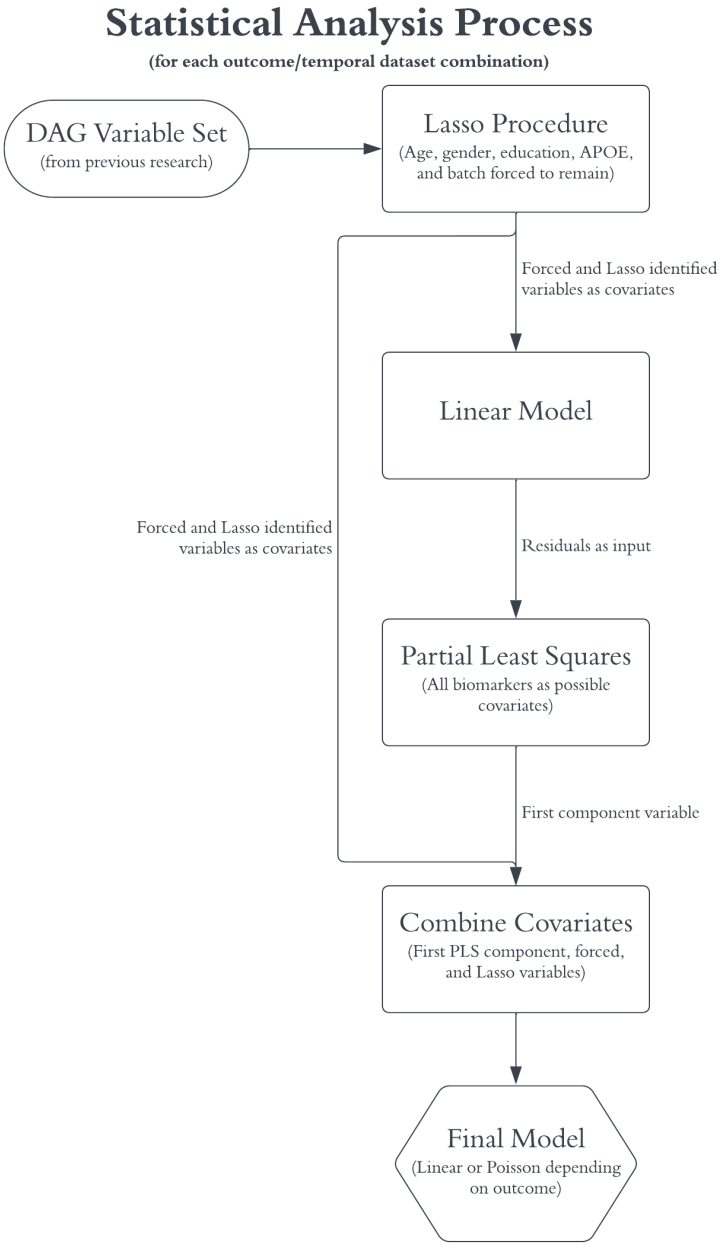
A major strength of this study was the availability of longitudinal data on well characterized research participants from an ADRC. Additionally, our sample size is comparable, if not larger, than those in the current similar literature [128,138-142]. We attempted to account for the complexity of interpreting data from multiple biomarkers by delving into methods beyond standard linear modeling. We were also able to investigate a relatively large number of biomarkers in our study compared to the published literature [105,107-108,112,128,140-141,143].

While our study sample represents a community-based cohort, it is highly specific in many ways. Our sample comes from a specific region of Kentucky (Central Kentucky), potentially limiting its applicability to individuals in other geographic and cultural regions. Given the convenience nature of the sample, and the restricted region from within which our sample comes, the external validity of our results is limited. Race was not included in the study because the vast majority of our participants are White, which could hinder the discovery of relationships relevant to the entire target population. In addition to lack of ethnoracial diversity, UK ADRC participants are much more highly educated than the general population of older adults.

We hope that future research can be informed by our negative findings. While the use of biomarkers via PLS identified components did not prove to be useful here, different methodology, sampling, populations, or use of other available blood biomarkers may lead to clinically useful results.

Tables and Figures

Figure 1. Statistical Analysis Flow Chart



Note: Flow chart outlining the statistical analysis process for each outcome-dataset combination. DAG: directed acyclic graph; PLS: partial least square; APOE: apolipoprotein e. Previous research refers to Chapter 2.

Table 1. Biomarker Subtypes

Subtype	Biomarker
Neurodegenerative/AD	Aβ40 Aβ42 Aβ42/40 t_tau t_tau/Aβ42 ptau181 ptau181/Aβ42 NfLight
Vascular	PlGF MMP9
Inflammatory	IL6 IL8 IL10 IL1β TNFα

Note: Biomarker subtypes. AD: Alzheimer's Disease. Aβ: amyloid beta; ptau: phosphorylated tau; t_tau: total tau; NfLight: Neurofilament light chain; TNFα: tumor necrosis factor alpha; PlGF: placental growth factor; MMP9: matrix metallopeptidase 9; IL: interleukin.

Table 2. Study Sample Characteristics

Variable	Statistic	N Missing
Batch (13/18)	193 (70)	0
Age* (years)	76.29 (6.56)	0
Gender (Female)	177 (64)	0
Race (White)	261 (94)	0
APOE (any e4)	94 (34)	2
Education* (years)	16.67 (2.65)	0
V1 Clinical Diagnosis (N)	239 (86)	0
V2 Clinical Diagnosis (N)	187 (66)	0
V1 Diabetes (y)	37 (13)	0
V2 Diabetes (y)	38 (14)	4
V1 Depression (y)	68 (25)	0
V2 Depression (y)	79 (28)	0
V1 Hypertension (y)	163 (59)	1
V2 Hypertension (y)	171 (62)	0

*Note: *Age and Education presented as mean(standard deviation), all other measures are N(%). APOE: apolipoprotein e; Clinical Diagnosis N = Normal. Diabetes, Depression, ang Hypertension coded as yes/no. V1: visit 1 information, V2: visit 2 information.*

Table 3. Lasso Identified Variables

Outcome	Dataset	Lasso Identified Variables		
CVLT-LD	1	-	-	-
	2	-	-	-
	3	-	-	-
	4	-	-	-
MMSE	1	Depression	-	-
	2	Depression	-	-
	3	Depression	-	-
	4	Depression	-	-
Trails B-A	1	-	-	-
	2	-	Diabetes	-
	3	-	-	Hypertension
	4	-	-	Hypertension

Note: Variables added to models via the Lasso procedure.
CVLD-LD: California Verbal Learning Test Long Delay; MMSE: Mini Mental State Exam; Trails: trail making test.

Table 4a. PLS First Component Coefficients

Outcome	Variable	Dataset 1 Coeff	Dataset 2 Coeff	Dataset 3 Coeff	Dataset 4 Coeff
CVLT-LD	AB40	0.0253	-0.0056	-0.0721	-0.0699
	AB42	0.0967	-0.0061	-0.0409	-0.0123
	AB42_40	-0.0090	-0.0368	0.0364	0.0583
	t_tau	0.0372	-0.0757	-0.1344	-0.1059
	ttau_ab42	-0.0562	-0.1307	-0.1420	-0.1192
	NfLight	-0.0474	-0.1847	-0.1831	-0.1180
	TNFa	-0.0494	-0.0284	0.0550	-0.1102
	PlGF	-0.0876	0.0375	0.1319	0.0579
	MMP9	0.0142	0.0143	-0.0619	-0.0716
	IL6	-0.0611	-0.0463	0.0863	0.0042
	IL8	-0.1839	-0.0362	-0.0020	0.0050
	IL10	-0.1502	0.1751	0.1433	0.0632
	IL1b	-0.0107	-0.0310	0.0840	0.0291
	ptau181	-0.0638	0.0380	0.1332	0.0903
	ptau181_ab42	-0.0300	-0.0070	0.0578	0.0714
	AB40 Diff	-	-	-	-0.0771
	AB42 Diff	-	-	-	-0.0593
	AB42_40 Diff	-	-	-	0.0374
	t_tau Diff	-	-	-	-0.0930
	ttau_ab42 Diff	-	-	-	0.0216
	NfLight Diff	-	-	-	-0.0343
	TNFa Diff	-	-	-	0.0265
	PlGF Diff	-	-	-	0.0392
	MMP9 Diff	-	-	-	-0.0451
	IL6 Diff	-	-	-	0.0180
	IL8 Diff	-	-	-	-0.0324
	IL10 Diff	-	-	-	-0.0165
	IL1b Diff	-	-	-	-0.0411
	ptau181 Diff	-	-	-	0.0582
	ptau181_ab42 Diff	-	-	-	0.0448

Table 4. (cont.) PLS First Component Coefficients

Outcome	Variable	Dataset 1 Coeff	Dataset 2 Coeff	Dataset 3 Coeff	Dataset 4 Coeff
MMSE	AB40	-0.1611	0.0575	0.1531	0.0206
	AB42	-0.1467	-0.1194	0.1851	0.1738
	AB42_40	-0.1050	-0.1279	0.1140	0.1923
	t_tau	-0.0106	0.0417	0.1201	0.0700
	ttau_ab42	-0.0388	-0.1703	0.0808	0.0330
	NfLight	0.1862	0.5438	1.3337	1.4221
	TNFa	0.1785	0.2174	-0.0270	0.1709
	PlGF	0.1253	0.0621	-0.1922	-0.2045
	MMP9	-0.0028	-0.0570	0.0742	0.1061
	IL6	0.1125	-0.0255	0.1652	0.0928
	IL8	0.0317	0.3038	0.3105	0.2765
	IL10	0.2483	0.8507	-0.1758	-0.2682
	IL1b	-0.0519	-0.0252	-0.2653	0.1165
	ptau181	-0.0176	0.2129	-0.2070	0.0004
	ptau181_ab42	-0.0421	-0.1315	-0.0153	0.1418
	AB40 Diff	-	-	-	-0.1548
	AB42 Diff	-	-	-	0.0143
	AB42_40 Diff	-	-	-	0.0476
	t_tau Diff	-	-	-	0.0537
	ttau_ab42 Diff	-	-	-	0.1679
	NfLight Diff	-	-	-	0.8240
	TNFa Diff	-	-	-	-0.1438
	PlGF Diff	-	-	-	-0.1028
	MMP9 Diff	-	-	-	0.0668
	IL6 Diff	-	-	-	0.1440
	IL8 Diff	-	-	-	-0.0966
	IL10 Diff	-	-	-	0.2216
	IL1b Diff	-	-	-	0.1395
	ptau181 Diff	-	-	-	0.0364
	ptau181_ab42 Diff	-	-	-	0.2836

Table 4. (cont.) PLS First Component Coefficients

Outcome	Variable	Dataset 1 Coeff	Dataset 2 Coeff	Dataset 3 Coeff	Dataset 4 Coeff
Trails B-A	AB40	-0.0120	0.0118	-0.0441	-0.0504
	AB42	0.0591	0.0435	0.0080	0.0255
	AB42_40	0.0715	0.0648	0.0569	0.0641
	t_tau	-0.0075	0.0204	0.0573	0.0507
	ttau_ab42	-0.0491	0.0063	0.0713	0.0546
	NfLight	0.0624	0.0325	0.0582	0.0704
	TNFa	0.0295	0.0448	0.0582	0.0747
	PlGF	-0.0169	0.0020	0.0034	-0.0097
	MMP9	-0.0237	-0.0355	0.0062	0.0485
	IL6	0.0563	-0.0265	0.0979	0.0743
	IL8	0.1217	0.0841	0.0732	0.0628
	IL10	-0.1010	-0.0260	-0.0210	0.0156
	IL1b	-0.0707	-0.0363	0.0514	-0.0227
	ptau181	-0.0030	-0.0250	0.0297	0.0201
	ptau181_ab42	-0.0013	-0.0772	0.0286	0.0402
	AB40 Diff	-	-	-	-0.0546
	AB42 Diff	-	-	-	-0.0070
	AB42_40 Diff	-	-	-	0.0538
	t_tau Diff	-	-	-	0.0428
	ttau_ab42 Diff	-	-	-	0.0048
	NfLight Diff	-	-	-	0.0154
	TNFa Diff	-	-	-	0.0633
	PlGF Diff	-	-	-	0.0172
	MMP9 Diff	-	-	-	0.0463
	IL6 Diff	-	-	-	0.0578
	IL8 Diff	-	-	-	-0.0205
	IL10 Diff	-	-	-	0.0249
	IL1b Diff	-	-	-	0.0038
	ptau181 Diff	-	-	-	-0.0031
	ptau181_ab42 Diff	-	-	-	0.0447

Note: Biomarker variable coefficients for first PLS components from each dataset.
CVLT-LD: California verbal learning test long delay; MMSE: Mini mental state exam; Trails: trail making test. Aβ: amyloid beta; ptau: phosphorylated tau; t_tau: total tau; NfLight: Neurofilament light chain; TNFα: tumor necrosis factor alpha; PlGF: placental growth factor; MMP9: matrix metallopeptidase 9; IL: interleukin. Coeff: coefficient; Diff: difference from study visit 1 to study visit 2.

Table 5. Final Model Results

Outcome	Variable	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
		Estimate	SE	P-values	Estimate	SE	P-values	Estimate	SE	P-values	Estimate	SE	P-values
CVLT-LD	Batch	0.064	0.242	0.792	-0.022	0.311	0.945	-0.181	0.338	0.594	-0.283	0.406	0.488
	Age	0.041	0.018	0.022	-0.025	0.024	0.296	-0.039	0.027	0.154	-0.020	0.031	0.519
	Gender	0.234	0.231	0.314	-0.186	0.286	0.518	-0.742	0.322	0.023	-0.549	0.402	0.177
	Education	0.088	0.039	0.025	0.081	0.049	0.102	0.115	0.055	0.041	0.092	0.066	0.168
	APOE	-0.332	0.238	0.165	0.134	0.292	0.646	-0.140	0.320	0.663	0.355	0.403	0.382
	PLS Component	-2.51E-06	3.55E-05	0.944	3.79E-06	3.95E-05	0.924	9.45E-06	1.15E-05	0.414	9.51E-06	8.36E-06	0.260
MMSE	Batch	0.052	0.150	0.731	0.195	0.103	0.060	0.145	0.127	0.252	-0.038	0.140	0.786
	Age	0.028	0.012	0.018	0.015	0.008	0.067	0.011	0.009	0.197	0.005	0.009	0.554
	Gender	-0.343	0.144	0.017	-0.187	0.102	0.068	-0.403	0.126	0.001	-0.343	0.139	0.014
	Education	-0.098	0.025	<0.001	-0.046	0.018	0.011	-0.101	0.021	<0.001	-0.075	0.022	0.001
	APOE	0.687	0.147	<0.001	0.939	0.103	<0.001	0.819	0.125	<0.001	0.912	0.140	<0.001
	Depression	0.498	0.159	0.002	0.461	0.112	<0.001	1.039	0.115	<0.001	0.953	0.126	<0.001
	PLS Component	1.57E-06	1.41E-04	0.991	2.24E-06	5.51E-06	0.684	5.32E-06	3.98E-06	0.181	2.29E-06	2.08E-06	0.271
Trails B-A	Batch	0.068	0.108	0.532	0.136	0.124	0.274	0.140	0.145	0.338	0.017	0.183	0.924
	Age	0.015	0.008	0.056	0.042	0.010	0.000	0.029	0.011	0.009	0.030	0.013	0.022
	Gender	-0.006	0.103	0.952	-0.095	0.120	0.428	-0.026	0.140	0.855	-0.050	0.176	0.776
	Education	-0.014	0.018	0.433	-0.020	0.021	0.336	-0.021	0.025	0.410	-3.410E-05	0.030	0.999
	APOE	0.145	0.107	0.174	0.192	0.124	0.123	0.286	0.144	0.049	0.256	0.182	0.164
	Diabetes	-	-	-	0.154	0.161	0.340	-	-	-	-	-	-
	Hypertension	-	-	-	-	-	-	0.122	0.138	0.378	0.184	0.172	0.290
	PLS Component	6.17E-06	9.62E-06	0.522	5.99E-06	7.24E-06	0.409	5.73E-06	4.81E-05	0.905	4.63E-06	4.23E-06	0.277

Note: Model results for all combinations of outcome and dataset, including prognostic variables, Lasso identified variables, and first PLS component (linear combination of biomarkers). CVLT-LD: California verbal learning test; MMSE: mini mental state exam; Trails: Trail making test. Batch represents the years 2013/2018. Age and education presented in years. Gender represents females, APOE is any e4 alleles. Depression, Diabetes, and Hypertension coded as ever having the condition. SE: standard error.
PLS components described in Table 3.

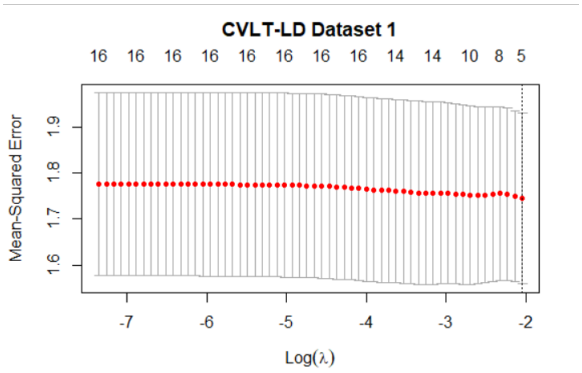
Table 6. Post Hoc Cognitive Outcome Model Results

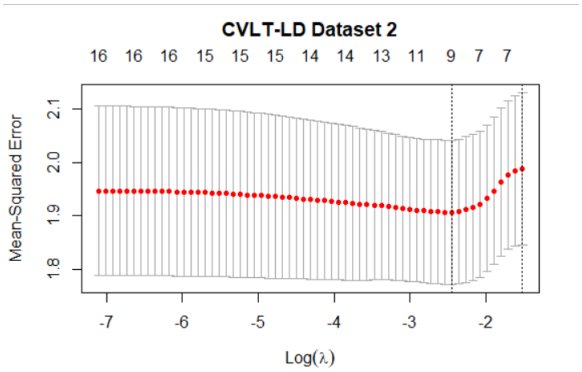
Outcome	Variables	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
		Estimate	SE	P-values	Estimate	SE	P-values	Estimate	SE	P-values	Estimate	SE	P-values
Cognition	Batch	0.028	0.065	0.67	0.077	0.079	0.334	0.082	0.09	0.363	0.003	0.111	0.978
	Age	-0.004	0.005	0.414	0.015	0.006	0.016	0.019	0.006	0.003	0.015	0.008	0.054
	Gender	-0.089	0.062	0.156	-0.268	0.076	0.001	-0.243	0.088	0.007	-0.268	0.109	0.016
	Education	-0.014	0.011	0.185	-0.023	0.013	0.086	-0.017	0.015	0.263	-0.011	0.018	0.528
	APOE	0.124	0.064	0.055	0.159	0.078	0.044	0.145	0.089	0.106	0.113	0.112	0.315
	Depression	-	-	-	0.305	0.086	0.001	0.321	0.089	0	0.373	0.106	0.001
	PLS Component	6.03E-06	7.21E-06	0.404	6.51E-06	6.13E-06	0.29	5.75E-06	9.61E-06	0.551	6.86E-06	4.87E-06	0.163

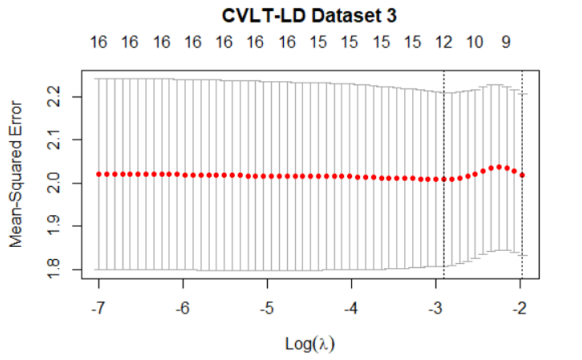
Note: Model results for post hoc analysis following the same analysis procedure with cognition as the outcome. Cognition is measured as Normal vs. any cognitive decline. Batch represents years 2013/2018, gender is female, APOE represents any e4 alleles, and depression represents ever having the condition. Age and education are measured in years. APOE: apolipoprotein e; PLS: partial least squares; SE: standard error.

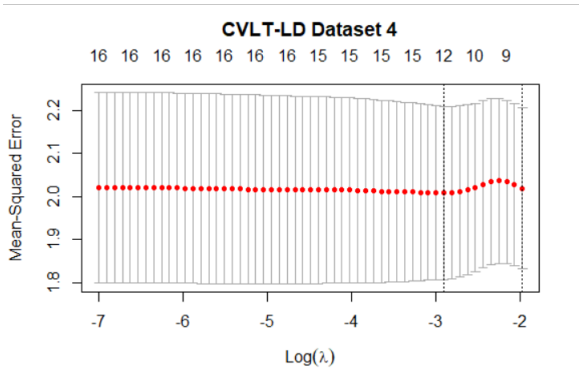
Supplemental Materials

Figure 1. Lasso Lambda MSE Plots (CVLT-LD)



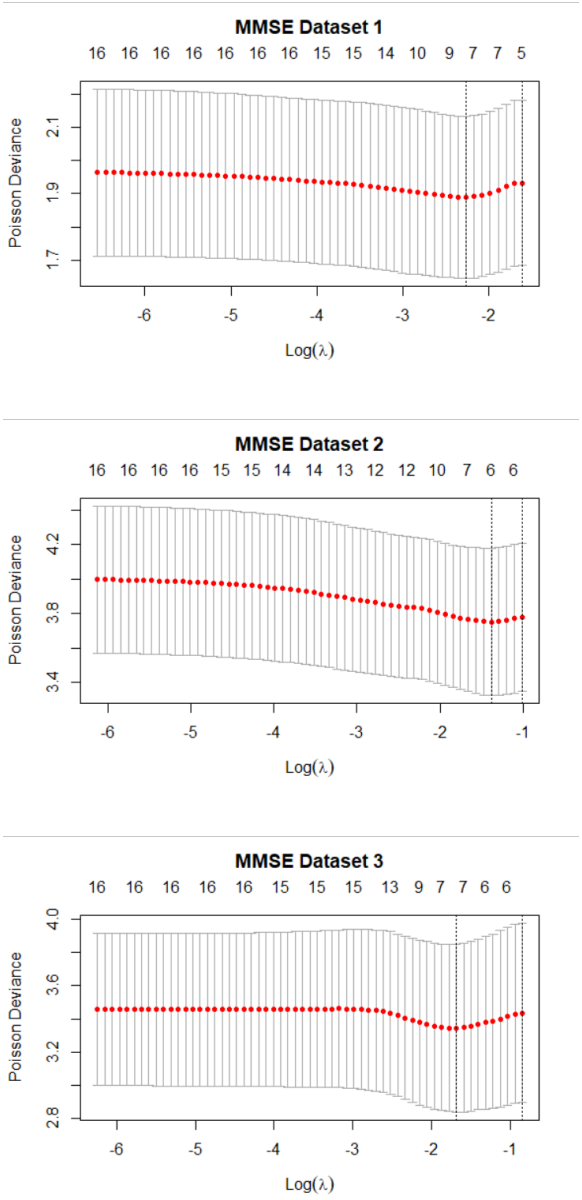


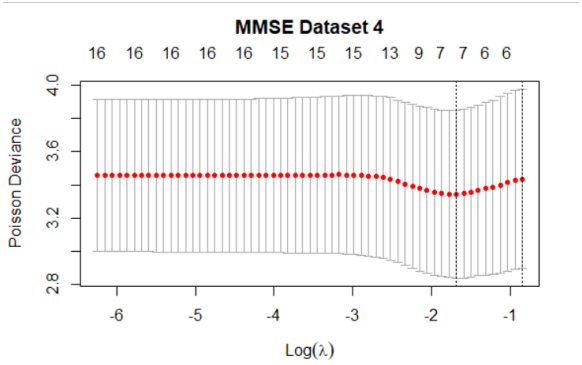




Note: Plots of mean squared error per model defined by different values of lambda from the Lasso procedure for each dataset with CVLD-LD outcome. Numbers below title represent number of variables left in model. Dotted vertical line for lowest value of MSE represents best lambda value. Remaining vertical dotted line represents second best lambda. CVLD-LD: California Verbal Learning Test Long Delay score; MSE: mean squared error.

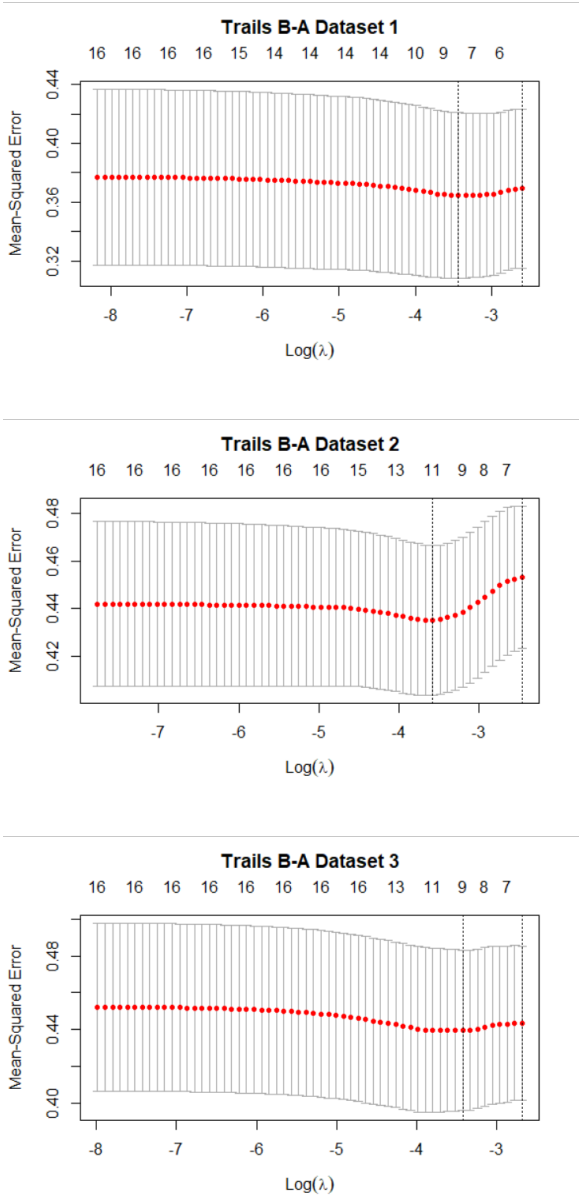
Figure 2. Lasso Lambda MSE Plots (MMSE)

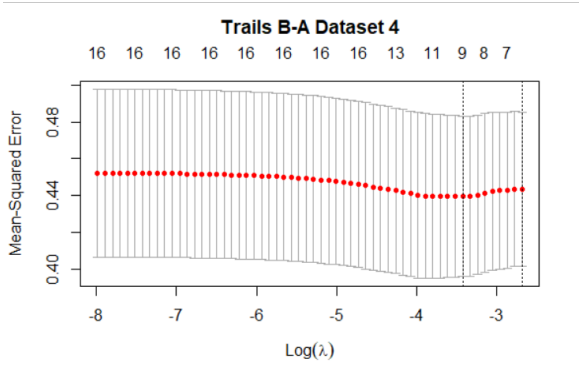




Note: Plots of mean squared error per model defined by different values of lambda from the Lasso procedure for each dataset with MMSE outcome. Numbers below title represent number of variables left in model. Dotted vertical line for lowest value of MSE represents best lambda value. Remaining vertical dotted line represents second best lambda. MMSE: Mini Mental State Exam; MSE: mean squared error.

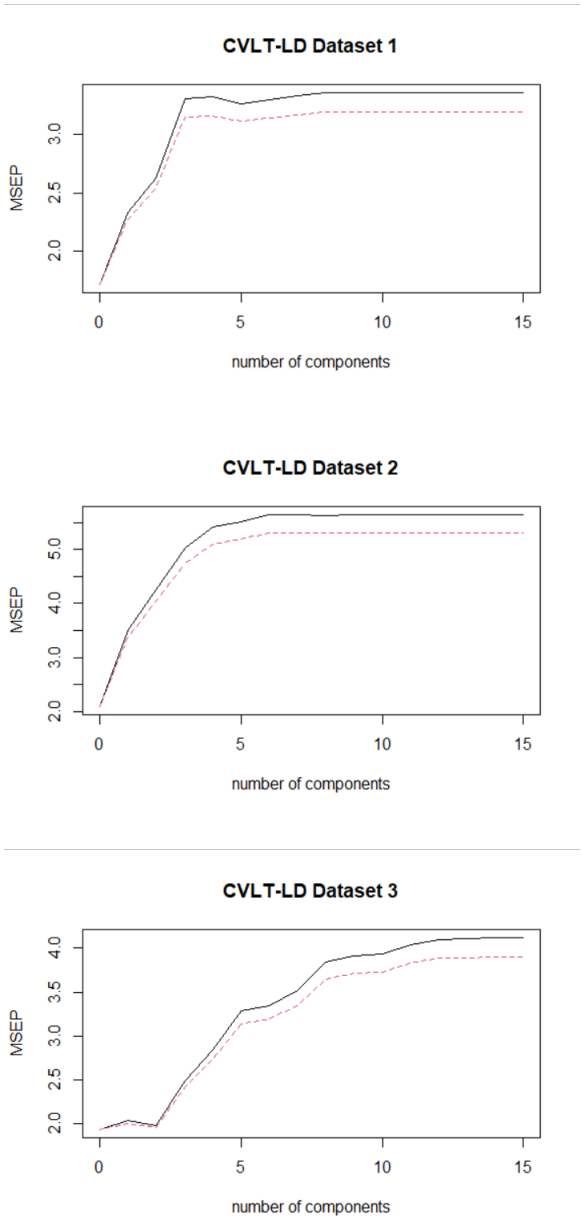
Figure 3. Lasso Lambda MSE Plots (Trails B-A)

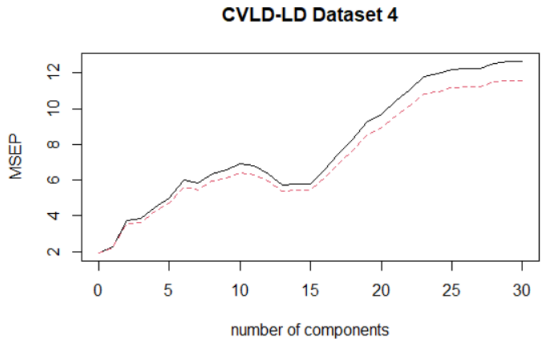




Note: Plots of mean squared error per model defined by different values of lambda from the Lasso procedure for each dataset with Trails B-A outcome. Numbers below title represent number of variables left in model. Dotted vertical line for lowest value of MSE represents best lambda value. Remaining vertical dotted line represents second best lambda. Trails: Trail Making Test; MSE: mean squared error.

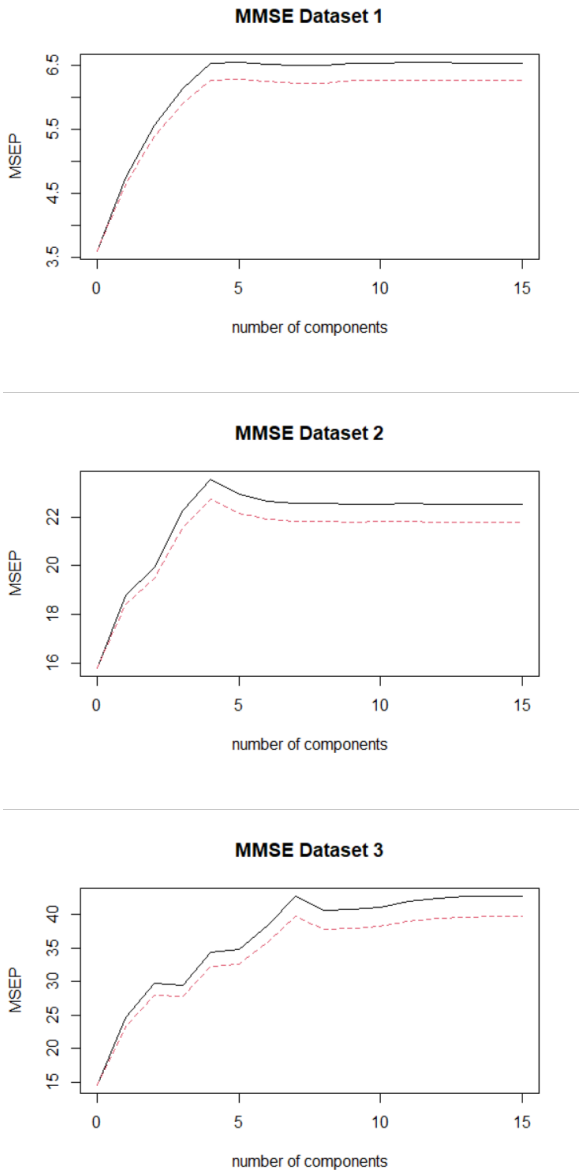
Figure 4. Partial Least Squares MSEP Plots (CVLT-LD)

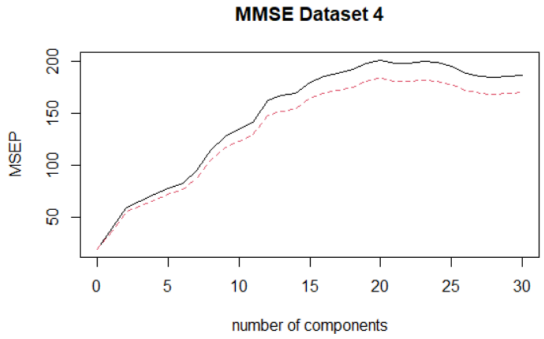




Note: Plots of mean squared error of prediction via cross-validation based on number of PLS components included, for each dataset with CVLD-LD outcome. MSEP: mean squared error of prediction; PLS: partial least squares; CVLD-LD: California Verbal Learning Test Long Delay.

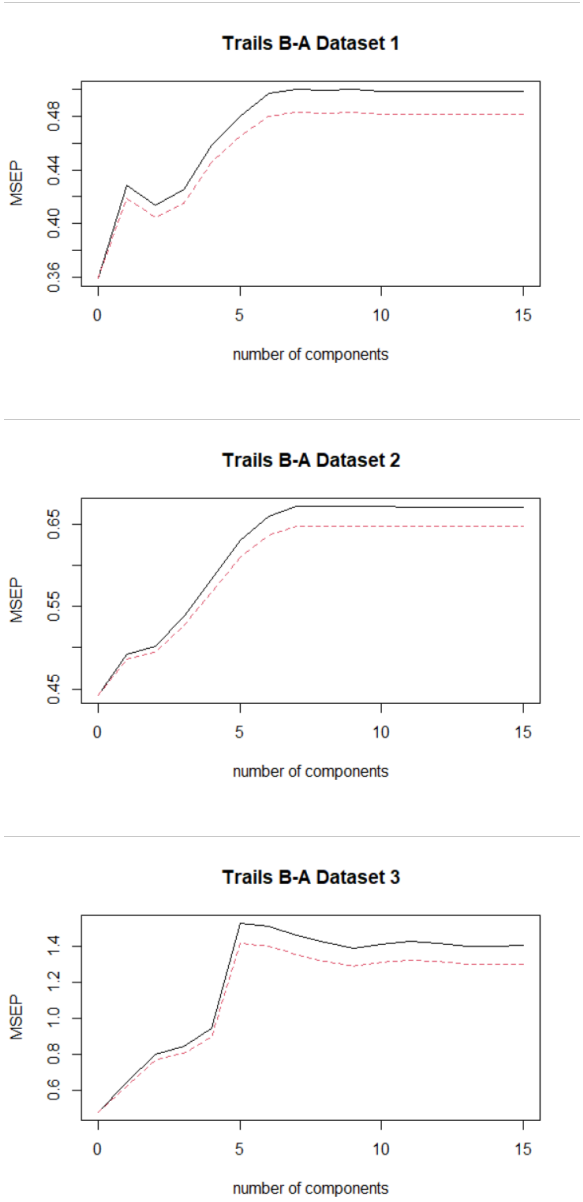
Figure 5. Partial Least Squares MSEP Plots (MMSE)

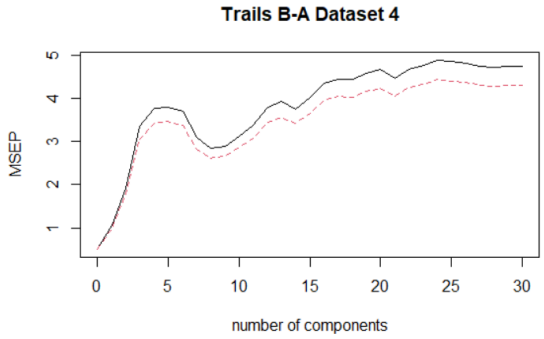




Note: Plots of mean squared error of prediction via cross-validation based on number of PLS components included, for each dataset with MMSE outcome. MSEP: mean squared error of prediction; PLS: partial least squares; MMSE: Mini Mental State Exam.

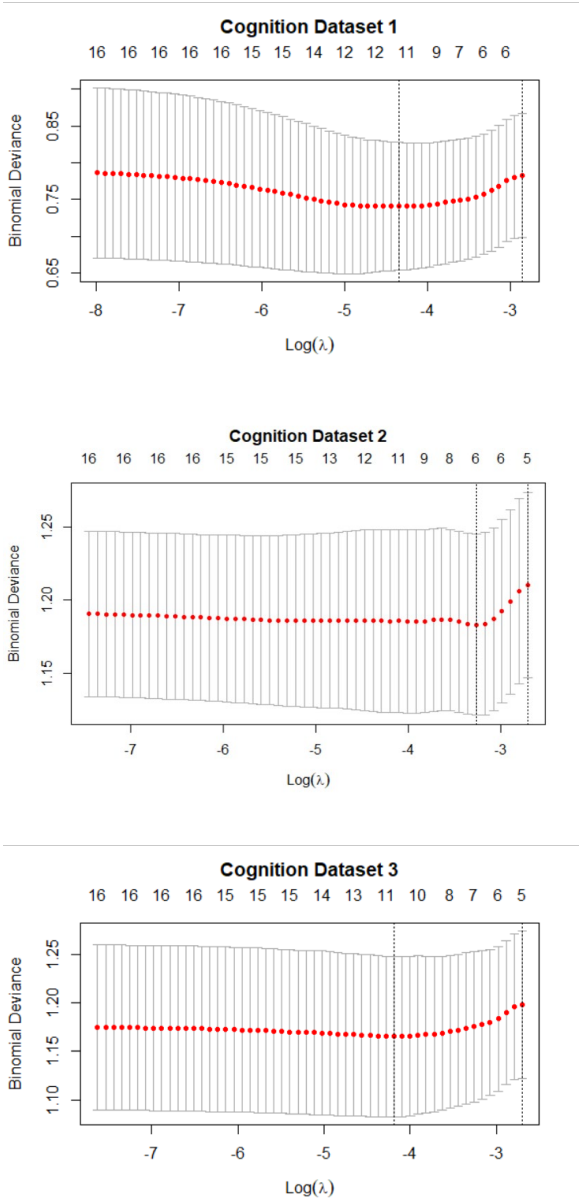
Figure 6. Partial Least Squares MSEP Plots (Trails B-A)

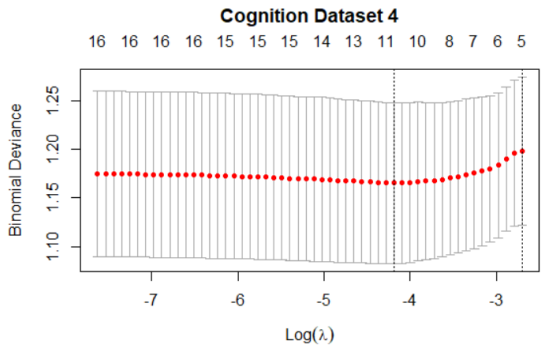




Note: Plots of mean squared error of prediction via cross-validation based on number of PLS components included, for each dataset with Trails B-A outcome. MSEP: mean squared error of prediction; PLS: partial least squares; Trails: Trail Making Test.

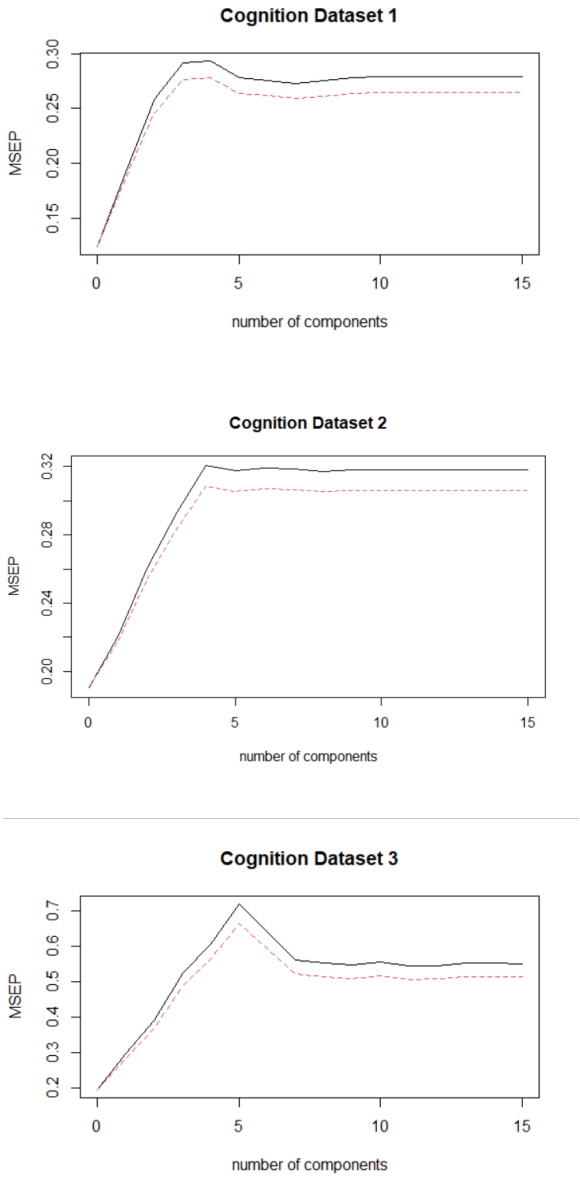
Figure 7. Post Hoc Lasso Lambda Binomial Deviance Plots (Cognition)

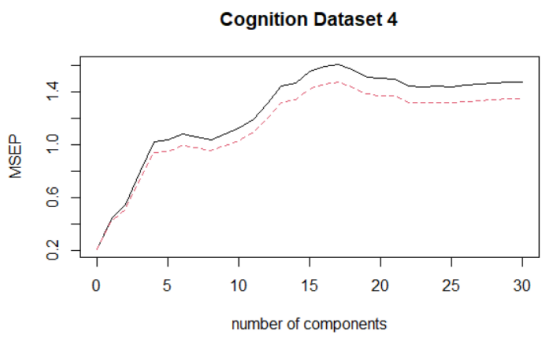




Note: Post hoc plots of binomial deviance per model defined by different values of lambda from the Lasso procedure for each dataset with cognitive diagnosis outcome. Numbers below title represent number of variables left in model. Dotted vertical line for lowest value of binomial deviance represents best lambda value. Remaining vertical dotted line represents second best lambda.

Figure 8. Post Hoc PLS MSEP Plots (Cognition)





Note: Post hoc plots of mean squared error of prediction via cross-validation based on number of PLS components included, for each dataset with cognitive diagnosis outcome. MSEP: mean squared error of prediction; PLS: partial least squares.

Table 1. Post Hoc PLS First Component Coefficients (Cognition)

Outcome	Variable	Dataset 1 Coeff	Dataset 2 Coeff	Dataset 3 Coeff	Dataset 4 Coeff
Cognition	AB40	0.0052	0.0188	0.0277	0.0148
	AB42	-0.0177	0.0017	0.0258	0.0254
	AB42_40	-0.0056	-0.0161	-0.0211	-0.0092
	t_tau	0.0057	-0.0080	0.0331	0.0248
	ttau_ab42	0.0428	0.0006	0.0352	0.0252
	NfLight	0.0382	0.0462	0.0712	0.0503
	TNFa	0.0164	-0.0060	0.0234	0.0332
	PlGF	0.0116	-0.0204	-0.0420	-0.0316
	MMP9	-0.0191	-0.0275	0.0197	0.0138
	IL6	-0.0057	-0.0091	0.0091	0.0066
	IL8	0.0576	0.0409	0.0499	0.0343
	IL10	0.0333	0.0224	-0.0058	-0.0061
	IL1b	-0.0135	-0.0232	-0.0249	-0.0240
	ptau181	-0.0068	-0.0142	-0.0171	-0.0131
	ptau181_ab42	-0.0081	-0.0289	0.0067	-0.0020
	AB40 Diff	-	-	-	-0.0160
	AB42 Diff	-	-	-	-0.0076
	AB42_40 Diff	-	-	-	-0.0034
	t_tau Diff	-	-	-	0.0263
	ttau_ab42 Diff	-	-	-	0.0256
	NfLight Diff	-	-	-	0.0134
	TNFa Diff	-	-	-	0.0027
	PlGF Diff	-	-	-	-0.0137
	MMP9 Diff	-	-	-	0.0327
	IL6 Diff	-	-	-	0.0097
	IL8 Diff	-	-	-	0.0134
	IL10 Diff	-	-	-	0.0099
	IL1b Diff	-	-	-	0.0206
	ptau181 Diff	-	-	-	0.0116
	ptau181_ab42 Diff	-	-	-	0.0256

Note: Post hoc biomarker variable coefficients for first PLS components from each dataset with cognition as the outcome. Cognition is presented as Normal vs. any cognitive decline.
Aβ: amyloid beta; ptau: phosphorylated tau; t_tau: total tau; NfLight: Neurofilament light chain; TNFα: tumor necrosis factor alpha; PlGF: placental growth factor; MMP9: matrix metallopeptidase 9; IL: interleukin.
Coeff: coefficient; Diff: difference from study visit 1 to study visit 2.

CHAPTER 4

Introduction

While prevention and cure are the end goal researchers work toward, clinical identification of Alzheimer's disease (AD) remains complicated as well. While certain identifiable pathologies are specific to AD, other pathological signs and symptoms are not AD specific and may overlap with other dementias and neurodegenerative disorders, making it difficult to differentiate [104,148-151]. The gold standard of AD diagnosis is still brain autopsy, but much focus is being given to the potential of blood-based biomarkers that can be measured throughout disease progression and before cognitive decline occurs. Diagnostic guidelines for AD-type dementia include biomarker evidence but recognize the limited validation of available biomarkers [65]. Thus, research focused on characterizing and validating blood-based biomarkers addresses a critical gap in knowledge.

Currently, hypothesized models of AD pathogenesis posit that measures of tau and amyloid beta, the hallmark proteins that are misfolded in AD neuropathological changes (ADNC), change gradually from normal to abnormal levels many years prior to the onset of cognitive symptoms [152]. Additionally, there is growing appreciation that AD-type dementia is most often explained by a mixture of ADNC and additional pathologies, including cerebrovascular disease and immune dysfunction [153-154]. However, longitudinal studies of blood-based biomarker trajectories are lacking.

Thus, in this study we examined age-related trajectories, via group-based trajectory models (GBTMs), of five plasma biomarkers that have shown promise in research related to cognitive decline: two neurodegenerative biomarkers, amyloid beta 42/40 (Aβ42/40) [155] and phosphorylated tau (ptau181) [156]; one vascular biomarker, placental growth factor (PlGF) [157]; and two inflammatory biomarkers, interleukin 6 (IL6) [158] and tumor necrosis factor alpha (TNFα) [159]. GBTM is a semiparametric method developed as an extension of mixture modeling [160-161]. Our goal was to identify distinct, latent trajectories for each of the biomarkers over time

and examine whether they differ based on cognitive performance. We hypothesized that if these biomarkers are truly measuring disease progression due to ADNC, then we would expect to see differences in cognitive performances between the GBTM identified trajectories.

Methods

Setting

Data for the current study were drawn from the University of Kentucky Alzheimer's Disease Research Center (UKADRC) community-based longitudinal cohort (full cohort description in [115]). Individuals were evaluated approximately annually until death and through autopsy, for which participants gave informed consent. Our sample was restricted to UKADRC participants who had plasma biomarker assays, which began collection in 2012. Due to technical changes in the acquisition of biomarker measurements (that is, switching from the Quanterix HD-1 analyzer to the HD-X analyzer), biomarker data from the first year of collection was excluded [115]. All studies have been reviewed and approved by the University of Kentucky Institutional Review Board (IRB).

Biomarkers

Blood-plasma biomarkers were measured using the Quanterix SiMOA HD-X machine, as described previously [115]. Quanterix multiplex kits were used to measure Aβ40 and Aβ42, while Quanterix uniplex kits were used to measure ptau181, PlGF, IL6, and TNFα. For analysis, values of Aβ42 were combined with Aβ40 to create the ratio of Aβ42/40. Aβ40 and Aβ42 were not investigated independently.

Risk factors

Participant age at each annual visit was rounded to the nearest year and used as the time scale for the GBTMs. Gender (male and female), education (years), Apolipoprotein E (*APOE*) (defined

as any e4 alleles vs no e4 alleles), and race (defined as white vs other) were included as baseline (non-time-dependent) risk factors in all models. While these baseline risk factors may aid in determining group membership, they are not time-varying and therefore have no effect on the shape of the trajectories [166].

Cognitive measures

Three measures of global cognition were used to evaluate the GBTMs: Mini-Mental State Exam (MMSE), Clinical Dementia Rating Global Score (CDRGLOB), and clinical cognitive diagnosis. The MMSE consists of 16 questions totaling 30 points [116]. MMSE scores are typically given as the number of correct answers out of 30, meaning the higher the score the better the cognition. While there are no universally agreed upon cutoffs for MMSE score interpretation, scores of 26 and above generally indicate normal cognition, while scores at or below 24 generally indicate cognitive impairment [162]. The MMSE is administered by trained psychometricians.

Clinical Dementia Rating Scale assessment yields a raw score called the CDR Sum of Boxes (CDRSUM; range 0-18, higher scores are worse) [163]. The CDR is administered by UKADRC clinicians and consists of six categories (memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care), each of which is rated by a study co-participant to reflect the participant's level of impairment: 0 (None), 0.5 (questionable), 1 (mild), 2 (moderate), or 3 (severe). Their ratings are added up to create CDRSUM [164]. An additional set of guidelines [163,165] is used to calculate a global CDR rating (CDRGLOB), which is a measure of overall severity of impairment presented on the same scale used to score the individual categories. Therefore, a lower global score suggests better cognitive function.

Finally, clinical cognitive diagnosis is assessed for each individual at each visit following a strict set of guidelines to reduce subjectivity across clinicians [70]. Diagnosis is based on physical examination, neurological examination, cognitive examination, and medical history, as well as

information provided by the study co-participant. Regression to a lesser clinical cognitive diagnosis following a more severe diagnosis was not allowed in this dataset.

Statistical Analysis

We used GBTMs to identify distinct trajectories for each of the five measured biomarkers. This method, developed by Nagin and colleagues [160-161,166-167], uses mixture modeling techniques to identify at least 2 latent subgroups within the population. These groups show how the outcomes progress over time [160-161,166-167]. This method is an effective tool for visually representing longitudinal changes in outcomes, and GBTMs have been increasingly used in recent years to model cognitive trajectories in ADRD research [168-171].

Just as the individual mixtures in classic mixture modeling do not represent literal groups [172], Nagin emphasizes that the identified trajectory groups should not be thought of as literal, identifiable groups of individuals, but rather as latent groups that follow similar courses over time [166]. Group membership is determined by maximum probability assignment and should be understood as probabilistic [166].

There are three different kinds of GBTMs: single, joint- (or dual-) trajectory, and multi-trajectory. We employed the use of both single and multi-trajectory models in this analysis. A single-trajectory analysis identifies the number of groups that develop similarly over time for one outcome. A dual-trajectory model is designed to analyze the interrelationship of two outcomes that jointly develop over time. The estimated trajectories for each outcome are linked with a table of conditional probabilities that measures the probability of following a specific trajectory for outcome 2 given that the individual is following a specific trajectory for outcome 1 [173]. In some cases, it may be desirable to allow trajectories to differ across multiple outcomes, but this method becomes exponentially computationally intensive when more than 2 outcomes are considered. Multi-trajectory models were designed to avoid the computational problem of the conditional probability tables when performed for more than 2 outcomes, but their use with just two outcomes

may be desirable as well. In this form of GBTM, the trajectories identified by the model are consistent across all outcomes being considered, meaning no matter how many outcomes are considered there is only one set of trajectory groups. This is much more computationally reasonable and also affords the researcher the knowledge that the individuals in a specific trajectory are the same across all outcomes used, which may be clinically meaningful.

For the current study, we fit multi-trajectory models to Aβ42/40 and ptau181 data (i.e., measures of ADNC), as well as IL6 and TNFα data (i.e., measures of inflammation), and we fit a single trajectory model to the PlGF data. In the initial runs of the GBTMs, there were many issues with convergence and calculation of standard errors. In order to combat these issues, two approaches were taken. First, all five biomarker outcomes were logged [174]. This not only reduced the range of values but allowed us to model the biomarker trajectories via the censored-Normal distribution within PROC TRAJ in SAS. As our outcomes were not censored, the minimum and maximum censoring values were set outside the range of logged values for each outcome, as advised by Nagin, for when outcomes are approximately Normal but not censored [167]. Second, the timescale for the models, age, was scaled and centered, as this increases model stability, as noted by Dr. Daniel Nagin and Dr. Bobby Jones [email correspondence, May 2023]. Once these measures were taken no more convergence or standard error issues were encountered.

Model selection for the GBTMs was undertaken as outlined in [175]. Prior to any modeling, the maximum reasonable number of groups per outcome was set to be 4. More than 4 trajectory groups may lead to additional convergence problems while also increasing difficulty of interpretation. With little known about the changes in biomarkers over time, and how those changes may relate to cognition, we believe 4 trajectories to be the maximum number of groups to consider here while remaining clinically relevant and interpretable. Again prior to any modeling, the maximum order for any one trajectory was set to be 2 (quadratic). In addition to being the recommendation by Nagin [166], we have no clinical reason to believe that the trajectory of biomarker measurements would be more complicated than can be described by a second order

polynomial. The original sample of eligible UKADRC participant visits was trimmed to include only those visits where the age at the visit was between 65 and 90, inclusive, as both tails were very scarcely populated [Figure 1].

The first step was to determine the number of trajectory groups for each of the 3 models. Since "the choice of the order of the trajectory for each group is of less importance than the choice of the number of such groups," [166(p67)], the order for each group in the group number selection process was set to quadratic and trajectory group orders were determined second. Trajectory group determination was done by comparing trajectory plots and Bayesian Information Criterion (BIC) for each set of models run with 1, 2, 3, and 4 groups, respectively. According to Nagin, the addition of a trajectory group is desirable only if the measured improvement of model fit, assigned to be the first component of the BIC, is larger than the incurred penalty for adding more parameters, the second component of the BIC [166(p65)], i.e., increasing the overall BIC by a measurable margin (as seen in Table 2 of Jones et. al. [167]), where twice the log of the Bayes factor can be approximated by twice the change in BIC between models and any difference greater than 6 supports the use of the more complex model via strong evidence.

Once the number of groups was determined, the orders of the polynomials describing the group mean of the trajectory over time were reduced from quadratic (where all orders were set for group number determination) to an appropriate level. This was done via manual backwards selection: within each model, each group was compared and the order coefficient with the highest p-value removed. This process was repeated until either the highest order coefficient per trajectory was statistically significant or only the 0-order remained. A 0-order polynomial (that is, a flat line) would suggest that the group mean for the trajectory of the outcome is stable over study time. A first order polynomial would suggest that the group mean for the trajectory of the outcome develops linearly (either positively or negatively) over study time. Finally, a second order polynomial for the mean of the trajectory would suggest that the outcome develops in a curvilinear fashion (with no changes in concavity, or points of inflection) over study time.

After group number and orders were determined, Wald tests [166(pp102-106)] and Likelihood Ratio tests [176] were performed for each of the 4 included baseline risk factors. Baseline risk factors, which do not change over time, are only useful as possible predictors of group membership. They are included throughout and tested for relevance at the end of the model selection process because they have "almost no impact on the form of the trajectories themselves," [166(p114)], because they don't contain information that influences the shape of the trajectory since they do not change over time [166(pp113-116)].

After the determination of the final models for each of the outcomes, the respective trajectories were compared in a myriad of ways. Average posterior probability for each assigned group was assessed to confirm each group surpassed the rule of thumb cutoff of 0.7, presented by Nagin [166(p88)]. Special consideration of trajectory groups must be taken when analyzing model results for the multi-trajectory models. As previously stated, in a multi-trajectory model, no matter how many outcomes are considered, there is only one set of groups, each containing the same set of individuals across all outcomes, though the trajectories for each group may differ across outcomes. Consequently, it may be the case that two (or more) trajectories look the same for one outcome, but some distinction between them can be found in a different outcome. In such a case, it may be beneficial to analyze trajectory characteristics with certain trajectories grouped together for one outcome when it appears that the distinction between them comes from the other outcome. When analyzing model results, final trajectory plots will be referenced for possible trajectory groupings that may provide clinically relevant results. Baseline risk factors (gender, *APOE*, education, and race) and cognitive assessments (MMSE, CDRGLOB, and clinical diagnosis) were tested for differences across assigned trajectory groups, and relevant groupings of trajectories, via ANOVA F tests [177] and Chi-square tests [178] for independence.

Post Hoc Analysis

A significant result for the Chi-Square test of independence of cognitive diagnosis transitions and trajectory group for Aβ42/40 and ptau181 trajectories led to a post hoc analysis further investigating the differences in cognition between the trajectory groups. First, we performed new Chi-Square tests of independence for baseline and final diagnoses separately. We also tested for differences in the number of individuals who transitioned to a more severe diagnosis, given it was possible for them to receive a more severe diagnosis. Additionally, we did pairwise Chi-Square comparisons for all combinations of trajectory groups for the three previously mentioned breakdowns of cognitive diagnosis to try and determine specifically which trajectories differed from the others.

Software

All results were interpreted at the 5% significance level. Analyses were performed using a combination of SAS version 9.4 and R version 4.1.2 in RStudio. The GBTM package was installed for use in SAS [167,179], while the following packages were used in R: tidyverse, plyr, haven, and readxl [86,136-137,180].

Results

After application of eligibility criteria, our sample included 1082 unique UKADRC participants, with an average of 5.3 study visits (range 1 to 10 visits). After trimming study visits where participants were younger than 65 or older than 90 years, our analysis sample comprised 1015 unique individuals with an average of 5.11 visits (range 1 to 10 visits). In all, 916 individuals were utilized (with complete information for the baseline risk factors, as required by PROC TRAJ in SAS). Among the individuals utilized in our analysis, approximately 60% were female, 87% were white, 39% had at least one *APOE* ε4 allele, and the average years of education attained was 16.25.

Model Selection

Each of the 3 GBTMs, two multi-trajectory and one single trajectory, resulted in a 4-group model. The 4-group model fit each outcome the best based on evaluation of trajectory plots [Figures 2-6] and BIC comparisons [Table 1], all of which exceeded the suggested cutoffs of 6 and 10 for justifying the larger model via strong and very strong evidence, respectively [167].

Two measures of BIC were used throughout both group number selection and order reduction for trajectory groups, one with N set to the number of observations and the other with N set to the number of unique individuals. Because N is intended to measure independent observations making up a sample, neither value of N defined here is technically correct in the GBTM setting because multiple observations from independent individuals are not themselves totally independent. The large and small values of N used to create these two BIC measures create a range, within which the theoretically correct BIC lies for each model [166(p68), 190], though this is debated [190]. Thus, comparisons were made using both values.

Order reduction for each of the trajectories was evaluated via manual backwards selection and comparison of model BICs. While the individuals assigned to a specific trajectory in a multi-trajectory model are consistent across outcomes, the order assigned to those trajectories can differ between outcomes. The orders for the first multi-trajectory model, for Aβ42/40 and ptau181, were reduced to final orders of 2, 2, 0, and 0 for Aβ42/40 and 1, 0, 0, 0 for ptau, for trajectories 1, 2, 3, and 4, respectively [Table 2]. The final orders for trajectories 1, 2, 3, and 4 for the single trajectory model for PlGF were 0, 0, 1, and 2, respectively [Table 3]. Lastly, the final orders for the second multi-trajectory model, for IL6 and TNFα, were reduced to orders of 2, 1, 1, and 0 for IL6, and 0, 0, 0, and 2 for TNFα, respectively [Table 4].

After the determination of the number and order of the trajectories for each model, baseline risk factors were assessed for their usefulness in group membership determination via commonly used LRTs [181] and Wald test (suggested in the GBTM literature [166,173]). Interestingly, results from the two versions of testing did not often agree on whether to retain the baseline risk factors.

Likelihood ratio tests for the inclusion of each risk factor for each model indicated to retain all variables except gender for Aβ42/40 and ptau, race for PlGF, and gender for IL6 and TNFα [Table 5], while the Wald tests suggested keeping only education and race for PlGF and the IL6 and TNFα models [Table 5]. In the interest of consistency, and in the light of these conflicting results, all four risk factors were retained for each GBTM.

Trajectory Characteristics

Following the determination of the final models for each GBTM, trajectory group characteristics, baseline risk factors, and cognitive performance were evaluated for each group and compared for differences. In total, every trajectory exceeded the 0.7 rule of thumb cutoff for average posterior probability set forth by Nagin [166(p88)], indicating adequate fit for each model.

In addition to evaluating cognitive characteristics across the assigned trajectories, the final trajectory plots were examined to determine any relevant groupings of trajectories for comparisons amongst trajectories used in multi-trajectory models. Evaluation of the Aβ42/40 plot suggests that trajectories 2, 3, and 4 may be similar and separate from trajectory 1 in relation to Aβ42/40, even though reasonable separation can be seen between them in relation to ptau181. Evaluation of the ptau181 plot suggests that trajectories 1 and 2 may be similar and separate from trajectories 3 and 4 in relation to ptau181, even though trajectories 1 and 2 appear very different in relation to Aβ42/40. This is the nature of multi-trajectory modeling, resulting from the requirement that individuals in each trajectory remain the same across outcomes, and thus we will be comparing relevant cognitive measures for the aforementioned groupings of Aβ42/40 and ptau181 trajectories [Figure 7], in addition to comparing each trajectory separately.

Similarly, evaluation of the final trajectory models for IL6 and TNFα revealed their own clinically relevant groupings for comparison. Evaluation of the final plot for IL6 suggests that trajectories 3 and 4 may be similar and separate from trajectories 1 and 2 in relation to IL6, even though trajectories 3 and 4 appear quite different in relation to TNFα. Evaluation of the final plot

for TNFα suggests that trajectories 1, 2, and 3 may be similar and separate from trajectory 4 in relation to TNFα, even though trajectories 1, 2, and 3 are distinguishable in relation to IL6. These clinically relevant groupings of trajectories for IL6 and TNFα were compared [Figure 9], again in addition to all 4 trajectories separately.

No groupings of trajectories were considered for the analysis of PlGF results because it was a single-trajectory model. The final PlGF plot confirmed that no trajectories should be combined for comparison [Figure 8].

Baseline Characteristics

Gender, race, and *APOE* status were compared across trajectory groups via Chi-square tests for independence, while education was assessed with an ANOVA F test. The ANOVA assumption of normality of education for the entire sample was tested via the Shapiro Wilk test [182] in R. Though the test failed and found evidence of non-Normality, an evaluation of the plot of years of education per individual suggests that we would not be egregious in proceeding with the analysis anyways [Supplemental Figure 1] as the F statistic is robust against non-Normality given the population is symmetric and unimodal, and group sizes are greater than 10 [183-184], both of which our data satisfy. Also tested was homogeneity of variances across groups for each set of trajectories, another assumption of the ANOVA test. For each set of trajectories, group variances were tested via Bartlett's test [185], and no evidence of unequal variances was found for any set of trajectories.

Among Aβ42/40 and ptau181 trajectories, each group closely resembled the total sample on gender and race. Consequently, the tests of independence found no significant difference between the groups. The distribution of *APOE* e4 alleles across trajectory groups differed slightly from the sample distribution for trajectory group 1, but statistical significance was not reached (possibly because group 1 was the smallest trajectory group). We were also unable to detect difference in the mean number of years of education across trajectory groups here [Table 6].

PlGF trajectory groups showed more differences in baseline characteristics than the Aβ42/40 and ptau181 trajectory groups. While group 2 appeared to resemble the total sample on gender, groups 3 and 4 appear to have a slightly higher percentage of males and group 1 was 100% female. The Chi-Square test for independence reached statistical significance, implying that the distribution of gender did differ across trajectory groups. For race, trajectory group 1 appears to have a vastly different distribution than the total sample (and the other trajectory groups), which was enough for the test of differences across groups to reach statistical significance. While there was some variation in *APOE* distribution across trajectory groups, they were not statistically significantly different. Finally, average education appeared to be lower than the sample average for group 1, and higher than the group average for group 3. The ANOVA test for differences in mean years of education across trajectory groups reached statistical significance [Table 7].

Lastly, unlike the Aβ42/40 and ptau 181 trajectory groups, TNFα and IL6 trajectory groups did show some differences in baseline characteristics across trajectory groups, though not as many as the PlGF trajectory groups. While trajectory group 3 appears to have a slightly higher percentage of females than the sample as a whole, the differences in gender distribution across trajectory groups were not enough to reach statistical significance. Groups 2, 3, and 4 appear to resemble the sample on race, but trajectory group 1 has a smaller proportion of white individuals. This difference was enough for the test of differences across groups to reach statistical significance. We again see some variation in *APOE* distribution across trajectory groups, but there is not enough evidence to lead to statistical significance. Finally, trajectory groups 3 and 4 appear to have slightly lower average years of education. The differences in education across trajectories were enough to reach statistical significance [Table 8].

Cognitive Characteristics

First, MMSE and CDRGLOB were compared across assigned trajectories, as well as relevant groupings of trajectories (informed by the final trajectory plots). MMSE was evaluated via the

difference in first and last score to assess how much an individual's score decreases (i.e., got worse) over their study time. The differences in the maximum (best) and last (theoretically worst) MMSE scores were also evaluated to account for any testing variation that may have occurred in their scores while remaining temporally consistent with disease progression. CDRGLOB was similarly evaluated via the difference in first and last scores to assess how much an individual's score increased (again, got worse) over their study time. The difference in first (theoretically best) and last (worst) CDRGLOB score was also evaluated to account for any testing variation that may have occurred in their scores while remaining temporally consistent.

In the total sample, the average difference in first and last MMSE scores was 1.12, with a standard deviation of 3.28, meaning that on average an individual's MMSE score will decrease by an estimated 1.12 points from their first to their last visit, though the large standard deviation suggests this is not a strong pattern within the entire sample. This measure reached statistical significance when testing for differences between trajectory groups for the following scenarios: all 4 trajectories separately and trajectories 1 and 2 vs. 3 vs. 4 for Aβ42/40 and ptau181 trajectories [Table 6], and trajectory 1 vs. 2 vs. 3 and 4 for IL6 and TNFα trajectories [Table 8]. For the whole sample, the average difference in maximum and last MMSE scores was 1.65, with a standard deviation of 3.15. This implies a slightly larger decrease in score between an individual's maximum and last MMSE score in our sample, though again there is large standard deviation. This measure reached significance only for the comparison of trajectory groups 1 and 2 vs. 3 vs. 4 for Aβ42/40 and ptau181 trajectories.

In the total sample, the average difference in first and last CDRGLOB scores was -0.13, with a standard deviation of 0.36, meaning that on average an individual's CDRGLOB score will increase by 0.13, but a large standard deviation suggests that this is not a strong tendency within our sample. This measure reached significance when testing for differences between trajectory groups for the comparison of all 4 groups for PlGF trajectories, the only cognitive evaluation across trajectory groups to reach significance [Table 7]. For the total sample, the average difference in

first and maximum CDRGLOB score was -0.16, with a standard deviation of 0.36. This measure did not reach significance when testing differences across trajectory groups for any outcome.

Finally, clinical cognitive diagnosis was examined. For the entire sample, approximately 71% of individuals were diagnosed as Normal at their first visit, while 16% were MCI and 14% had dementia [Table 6]. The majority of baseline Normal individuals remained Normal at their last visit (~80%), while some transitioned to MCI (15%), and few reached Demented (5%) by their last visit. Approximately 76% of individuals with MCI at their first visit remained there at their last visit, while 24% progressed to Dementia. Given the restriction of no milder diagnoses after receiving a more severe diagnosis, all patients who had dementia at their first visit remained Demented at their last visit.

The diagnosis transition from first visit to last visit was compared across trajectories via Chi-square test. For Aβ42/40 and ptau181 trajectories it appears that trajectory 1 (which is by far the smallest group) may have a different distribution of diagnosis transitions than the rest of the groups [Table 6]. Though the Chi-square test does not specify which trajectories differ from the others, it did reach statistical significance, implying that diagnosis transition was not independent of trajectory group. Diagnosis transitions in the PlGF trajectory groups do not appear starkly different aside from trajectory group 1 (which is again by far the smallest group) having no individuals that were MCI at the first visit [Table 7]. The test for differences of diagnosis transitions across trajectory groups did not reach statistical significance. Finally, the distribution of diagnosis transitions across trajectory groups for TNFα and IL6 trajectories appear to be distributed roughly similarly, which is supported by a non-significant result from the Chi-square test for independence [Table 8].

Post Hoc Analysis

Following a significant Chi-Square test for diagnosis transitions for Aβ42/40 and ptau181 trajectories, further analysis was done to investigate three cognitive characteristics (baseline

diagnosis, final diagnosis, and diagnosis progression) across trajectory groups. The distribution of baseline cognitive diagnosis in trajectory groups 2, 3, and 4 resemble the distribution of the entire sample, but trajectory group 1 has far fewer Normal individuals at baseline. The distribution of final cognitive diagnosis for the whole sample is reflected in trajectory groups 2 and 4, but trajectory group 3 appears to have slightly more Normal individuals and trajectory group 1 has far fewer Normal individuals. Chi-Square tests concluded that both baseline and final cognitive diagnosis differed among trajectory groups [Table 9]. The final overall comparison made was a dichotomy of whether individuals in the sample progressed to a more severe diagnosis from their baseline to final diagnoses (amongst only those individuals who could progress, i.e., excluding individuals who began as Demented). While trajectory groups 2 and 4 appear similar to the total sample, groups 1 and 3 show lower proportions of individuals who did progress to a more severe diagnosis. This cognitive characteristic also proved to be significantly different across trajectory groups via a Chi-Square test [Table 9]. All pairwise combinations of trajectories were then compared via Chi-Square test for all three aforementioned cognitive characteristics. For both baseline and final cognitive diagnosis, trajectory group 1 was significantly different than groups 2, 3, and 4, but no significant differences could be found between trajectory groups 2, 3, and 4. For cognitive diagnosis progression, the only pair of trajectories that were statistically significantly different were trajectory groups 3 and 4 [Table 10].

For all three cognitive characteristics, pairwise Chi-Square comparisons with trajectory group 1 are not stable because >20% of cells have an expected cell size of <5. This is a result of the small size of trajectory group 1 and should be considered when weighing the validity of these results.

Discussion

Using data from a community-based cohort of older adult research volunteers, we used GBTMs to identify latent groups (trajectories) of five clinically relevant blood-plasma biomarkers,

investigating gender, race, education, and *APOE* as risk factors that influence trajectory group membership. We were able to attain adequate fit for all 3 models and compare baseline risk factors and cognitive assessments across trajectories for each outcome.

Baseline risk factors were evaluated for utility in predicting group membership by both Wald tests and LRTs. Test results for their inclusion from each method did not often agree. All four baseline risk factors were retained in each model because they each achieved significance for at least one model, via at least one testing method, and we wished to remain consistent across models. The source of the inconsistency of test results per risk factor likely stems from the difference in basis of each test. An LRT is based on a simple difference in log likelihoods [186] and is thought to be very reliable [181]. A Wald test, however, is based on the inverse of a Fisher information matrix [187], which is itself based on derivatives of the log likelihood function [187], the calculations of which may not be as stable, yet the test was recommended [166].

Interesting patterns are seen in the trajectories for each biomarker. While the majority of individuals have a stable pattern of development of Aβ42/40 over time, a small portion of individuals very much differ with a strong decrease in the ratio over time. These individuals are repeatedly highlighted as different in the analysis. The small size of this group of individuals suggests caution, but it is possible that such a subset of individuals exists in the population. The trajectory plot for ptau181 differs greatly from that of Aβ42/40. This plot suggests that the general level of ptau181 is more important and generally stable over time. PlGF, the only single-trajectory model where we would expect all trajectories to be distinguishable, shows that the general level of PlGF over time is likely its most important feature (though two of the trajectories were beyond 0-order). Similar to the first model, group 1 for PlGF is much smaller than the rest, which can account for the deviation of the nonparametric fit around the polynomial mean line. IL6 trajectories suggest that there are important differences in IL6 level, while all slightly increase over time. Throughout our trajectory analyses there are groups that are defined for subsets of the available ages of the sample individuals. This is a result of the overall age distribution and the fact that our individuals

cover a range of number of visits to be included in the data. TNFα has the trajectory defined for the shortest span of ages in its group 4, which is also its smallest group at N=28 individuals. Though this trajectory does not seem biologically plausible in relation to the rest of the individuals, its presence was strong in the data, and likely just a feature of our sample. These individuals were assigned their own trajectory when considering any number of trajectories beyond 1 [Figure 6]. However, as discussed below, although race and education are found to differ across trajectory groups, the distributions in group 4 do not immediately stand out as different than that of the entire sample for these or the non-significant baseline risk factors. Additionally, this trajectory group is not found to be significantly different from the rest in any cognitive characteristics.

Aβ42/40 and ptau181 trajectories were not found to differ on any baseline risk factors. The insignificant findings are perhaps not surprising given the disagreement between their respective LRT and Wald tests. What is surprising is the trajectory groupings for which significant MMSE scores were found when compared to the significant clinical diagnosis transitions and subsequent pairwise trajectory comparisons. Although the post hoc clinical diagnosis findings suggest that trajectory group 1 is significantly different from the other 3 trajectories, the grouping of trajectory 1 vs all three other trajectories were the only tests in which MMSE was not significant. The significant findings for MMSE are likely driven more by ptau181 than Aβ42/40 because the trajectory group pairings that reached significance represent the possible relevant pairings based on the final ptau181 trajectory plot [Figure 7]. However, significant cognitive diagnosis findings are perhaps driven more by Aβ42/40 because they mainly highlight the difference of trajectory group 1 from the other 3, which is reflected in the final trajectory plot for Aβ42/40 [Figure 7]. These findings show that there is some direct relation between neurodegenerative biomarkers and cognitive characteristics that extends beyond any relation they may have with the baseline risk factors. Based on the significant findings and the patterns seen in the final Aβ42/40 and ptau181 plots, it appears that, in general, level of ptau181 over time is important, while changes in the balance of Aβ42 and Aβ40 may be influential.

Unlike the first model, the baseline gender, race, and education factors did differ across PlGF trajectory groups, though useful baseline risk factors did not appear to influence findings in cognitive characteristics across trajectory groups. Similar to the first model, it appears as though trajectory group 1 stands out from the other 3 in most measures, although there was not probable cause to investigate this difference further as the final trajectory plot did not prompt pairings of trajectories to compare (as is expected for a single-trajectory model) [Figure 8], and the test of independence of clinical diagnosis transitions across trajectory groups was not significant. PlGF trajectory groups were the only ones to show a significant difference for any CDRGLOB measure, perhaps suggesting vascular biomarkers may be more related to this measure than inflammatory or neurodegenerative biomarkers.

Although there appears to be an unreasonable trajectory defined for TNFα, this is likely just a feature of the sample [188], not the population, and the remainder of the model should still be thought of as statistically sound. Few significant results were found for the IL6 and TNFα trajectory groups, similar to that of PlGF. Race and education differed across groups, while gender and *APOE* did not. The only significant finding in cognitive characteristics was the difference of first and last MMSE score for the comparison of trajectory groups 1 vs. 2 vs. 3 and 4. This result is likely driven by IL6 because that trajectory group pairing is reflected in the final IL6 trajectory plot [Figure 9], although the same grouping of trajectories did not yield a significant finding for the difference of maximum and last MMSE scores. Unlike the previous two models, no trajectory groups here consistently stood out as possibly different from the rest, and no more significant differences were found in the cognitive characteristics.

This study utilized a relatively large number of individuals and took full advantage of their longitudinal data without loss of information due to breaking it down for a cross-sectional analysis. Sophisticated modeling techniques enabled the evaluation of complex relationships between biomarkers and cognition for a considerable range of ages relevant to disease progression. Additionally, the ability to examine five plasma biomarkers that cover a wide range of functions

within the body sets our study apart from the majority that examine very few biomarkers [107-110,128,140-141].

While the strength and value of this study is considerable, there are limitations to be acknowledged. Our sample size is quite reasonable, though limited to complete cases. The contribution of anywhere from 1 to 10 visits per individual means that we are unable to estimate these trajectories with the same amount of information at each age. Less information in the age extremes leads to less certainty or trajectories defined over only a portion of the study time (as seen in the final trajectory plots). Full information from individuals across the ages examined as study time could help fill out the ends of the trajectories and possibly rectify any odd trajectory shapes. There were many hypothesis tests performed in this analysis, especially with the addition of the post hoc analyses. No multiple testing corrections were added to any p-values in the post hoc analysis, so we must be wary of Type 1 errors. Additionally, many of the Chi-Square tests performed in the post hoc analysis were not stable because of small cell sizes, casting some level of doubt on their validity.

An important factor to consider when evaluating this study is the nature of the data. As previously stated, our sample includes multiple observations for individuals, ranging from 1 to 10 visits and averaging about 5 visits. This implies that there are individuals whose first, last, and maximum values all come from the same visit. These individuals would also be categorized as not progressing to a more severe diagnosis by their last visit (if they began as either Normal or MCI), even though they theoretically could not progress with only one visit. These individuals were included so we could gain as much information as possible for all factors (baseline, cognitive, and biomarker). Approximately 22% of the 916 individuals utilized in the analysis had only one visit. This represents a sizeable portion of the data and may have influenced the results. In the future it could be useful to evaluate data with and without individuals who have only one time point (or age in this case) to evaluate whether this feature has an effect on the analysis and eliminate it if necessary.

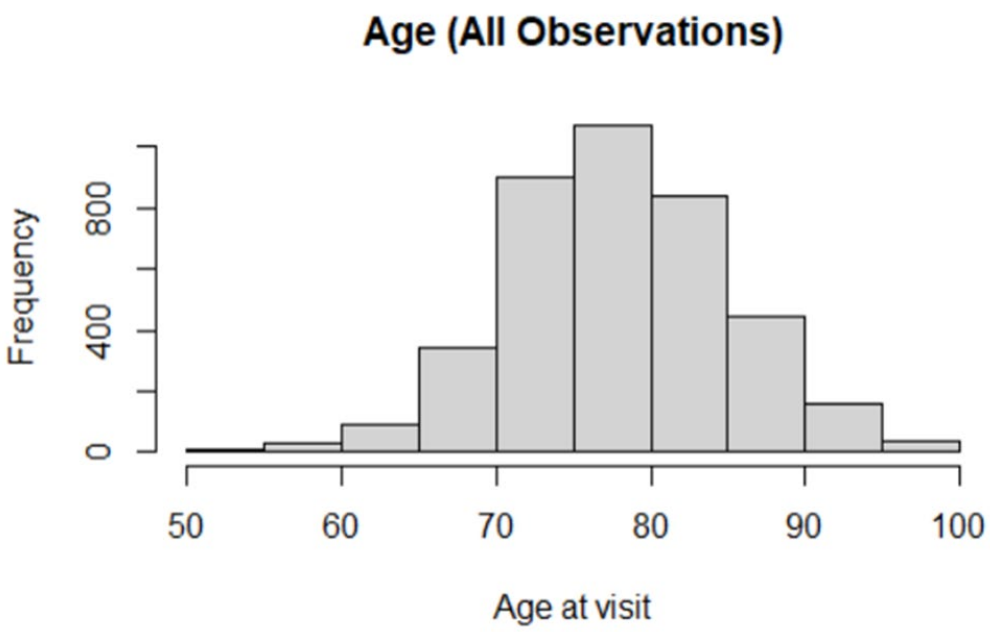
There are many other future directions researchers could explore, informed by our findings. Most obviously, it would be beneficial to similarly study additional plasma biomarkers to establish more relationships with cognitive characteristics over time. It may also be beneficial to examine some of our results, specifically the post hoc pairwise comparisons that had many small cell sizes, with different statistical tests that account for such limitations, like the non-parametric Fisher's exact test [178].

Given the significant results for AB42/40 and ptau181, it may be useful to examine these biomarkers separately, or even with a dual-trajectory model, to see if the positive results are driven by one or both of them. Also, we would like to perform a more in-depth examination of the changes in Aβ40 and Aβ42 over time, as new insights could be gained by examining them separately (outside of their ratio). While interpretation and computational difficulties may present new challenges, a dual-trajectory analysis of Aβ40 and Aβ42 could be an effective method because the trajectory groups, and even number of groups, would not be required to be the same across the two outcomes.

This study adds valuable information to the field of AD blood-plasma biomarker research with the use of a sophisticated statistical modeling technique to characterize biomarker trajectories that develop over many years. Our findings suggest that Aβ42/40 and ptau181 are related to MMSE measures and clinical cognitive diagnosis over time, and that PlGF and IL6 may be related to CDRGLOB and MMSE measures over time, respectively. These findings highlight important features of the relation of two neurodegenerative biomarkers to cognitive measures over time and provide a promising basis for future research of these elements with more neurodegenerative, vascular, and inflammatory biomarkers.

Tables and Figures

Figure 1. Age Histogram



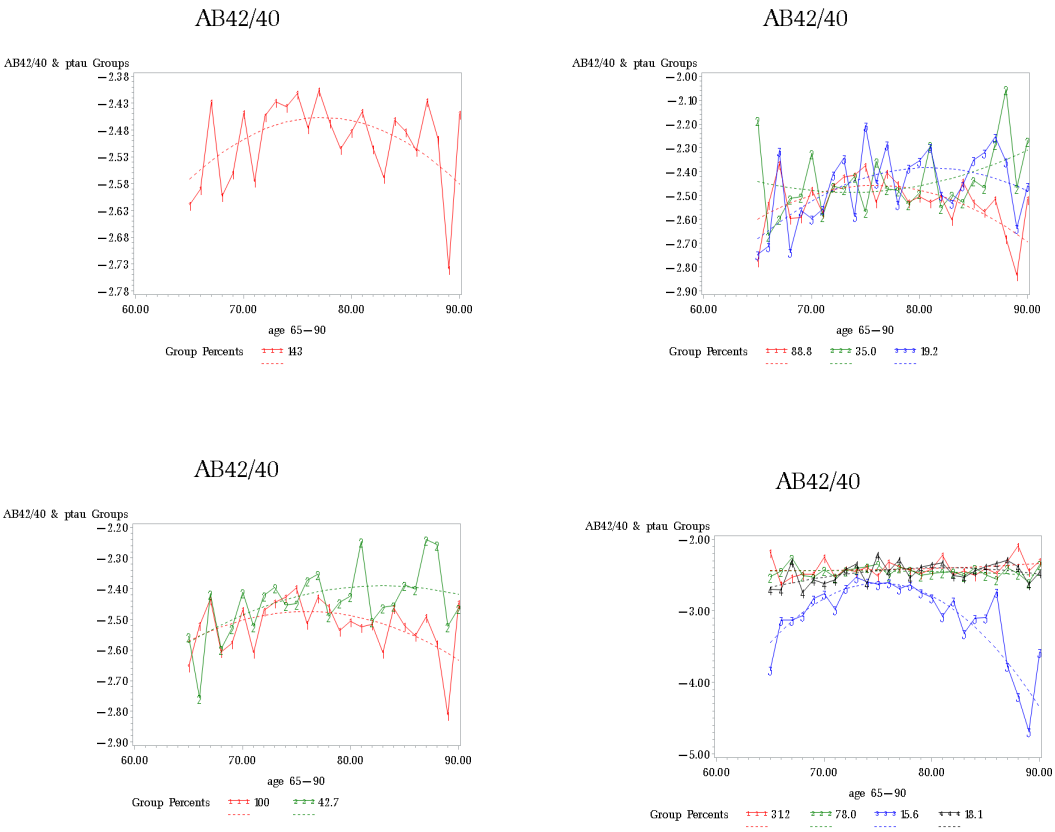
NOTE: Histogram of age for each visit originally available for analysis. Multiple observations per individual included. Age recorded in years.

Table 1. Model Group Number Selection Statistics

Outcome(s)	Num. of groups	Big N BIC	(Big N)	Small N BIC	(Small N)	Group Comparison	Big BIC Diff	Small BIC Diff	2*Big Diff	2*Small Diff
Aβ42/40 + ptau	1	-6809.85	4900	-6801.71	641	-	-	-	-	
	2	-6314.29	4900	-6294.97	641	2 - 1	495.56	506.74	991.12	1013.48
	3	-6212.77	4900	-6182.26	641	3 - 2	101.52	112.71	203.04	225.42
	4	-6182.98	4900	-6141.28	641	4 - 3	29.79	40.98	59.58	81.96
PlGF	1	-5343.04	3207	-5340.64	965	-	-	-	-	-
	2	-4286.41	3007	-4278.78	843	2 - 1	1056.63	1061.86	2113.26	2123.72
	3	-4180.12	3007	-4167.4	843	3 - 2	106.29	111.38	212.58	222.76
	4	-4175.96	3007	-4158.16	843	4 - 3	4.16	9.24	8.32	18.48
IL6 + TNFα	1	-9830.46	6571	-9822.4	879	-	-	-	-	-
	2	-9370.16	6571	-9351.02	876	2 - 1	460.3	471.38	920.6	942.76
	3	-9226.29	6571	-9196.06	876	3 - 2	143.87	154.96	287.74	309.92
	4	-9198.13	6571	-9156.82	876	4 - 3	28.16	39.24	56.32	78.48

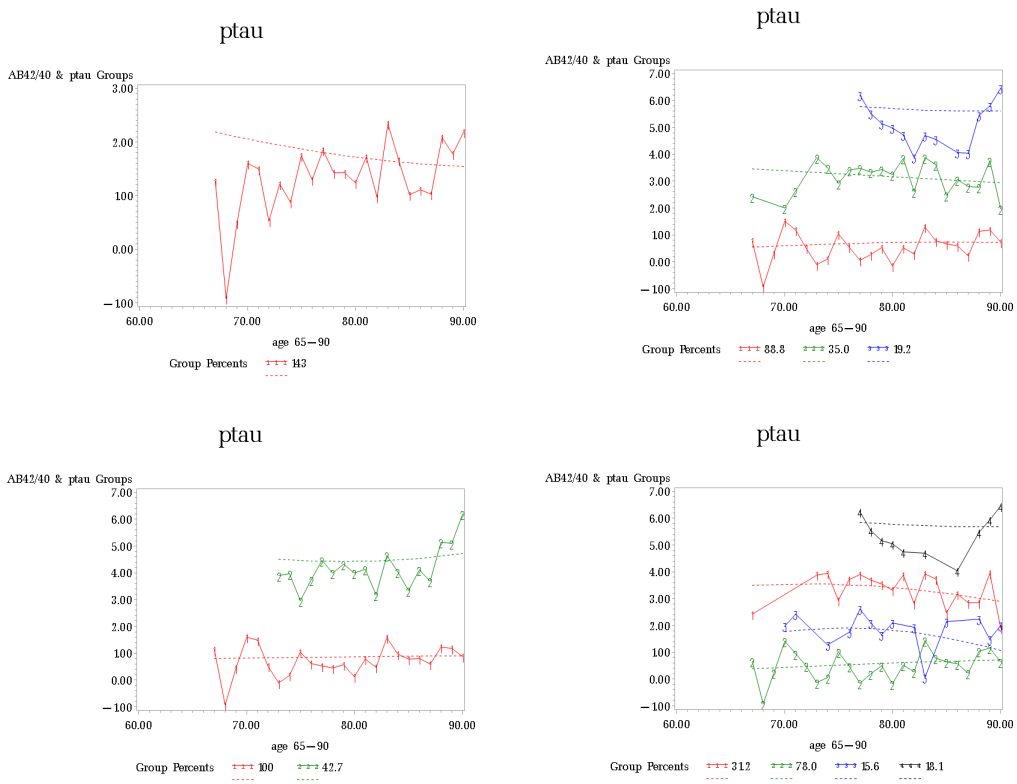
NOTE: Sample sizes are based on available biomarker and risk factor information per individual.
Diff: Difference; BIC: Bayesian information criterion; Aβ: amyloid beta; PlGF: placental growth factor; IL: interleukin; TNFα: tumor necrosis factor alpha.

Figure 2. Aβ42/40 Group Number Selection Plots



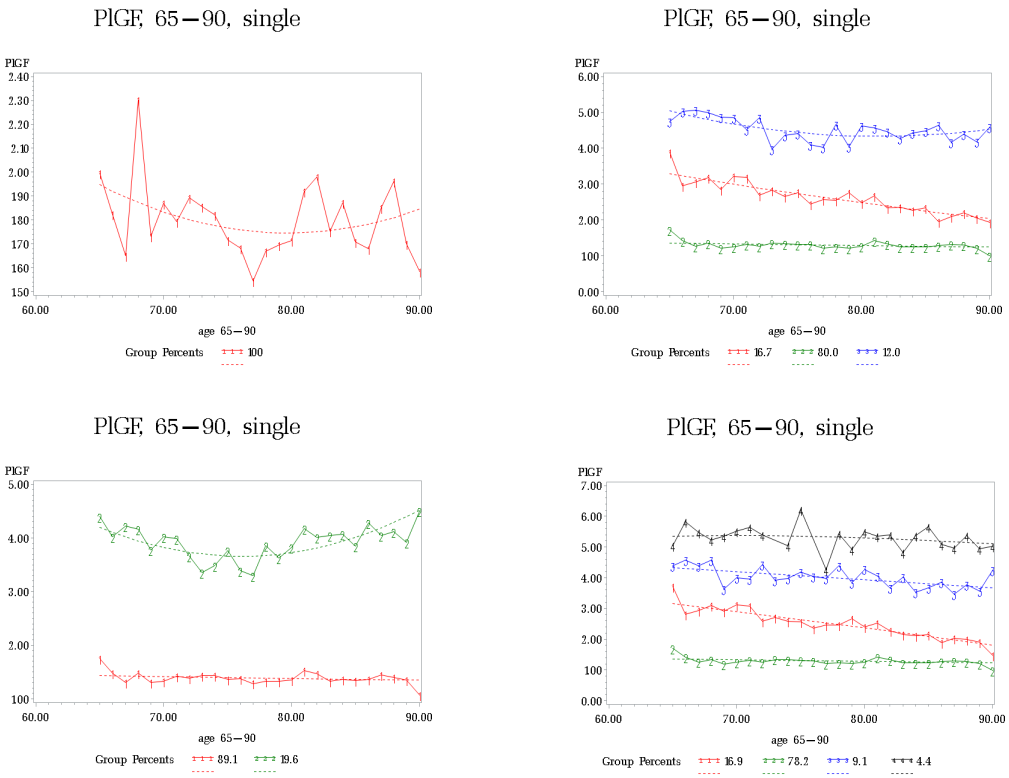
NOTE: Aβ42/40 trajectory plots based on multi-trajectory GBTMs for the outcomes Aβ42/40 and ptau181 with 1, 2, 3, and 4 trajectory groups considered. GBTM: group-based trajectory model; Aβ: amyloid beta; ptau: phosphorylated tau.

Figure 3. Ptau181 Group Number Selection Plots



NOTE: Ptau181 trajectory plots based on multi-trajectory GBTMs for the outcomes Aβ42/40 and ptau181 with 1, 2, 3, and 4 trajectory groups considered. GBTM: group-based trajectory model; Aβ: amyloid beta; ptau: phosphorylated tau.

Figure 4. PlGF Group Number Selection Plots

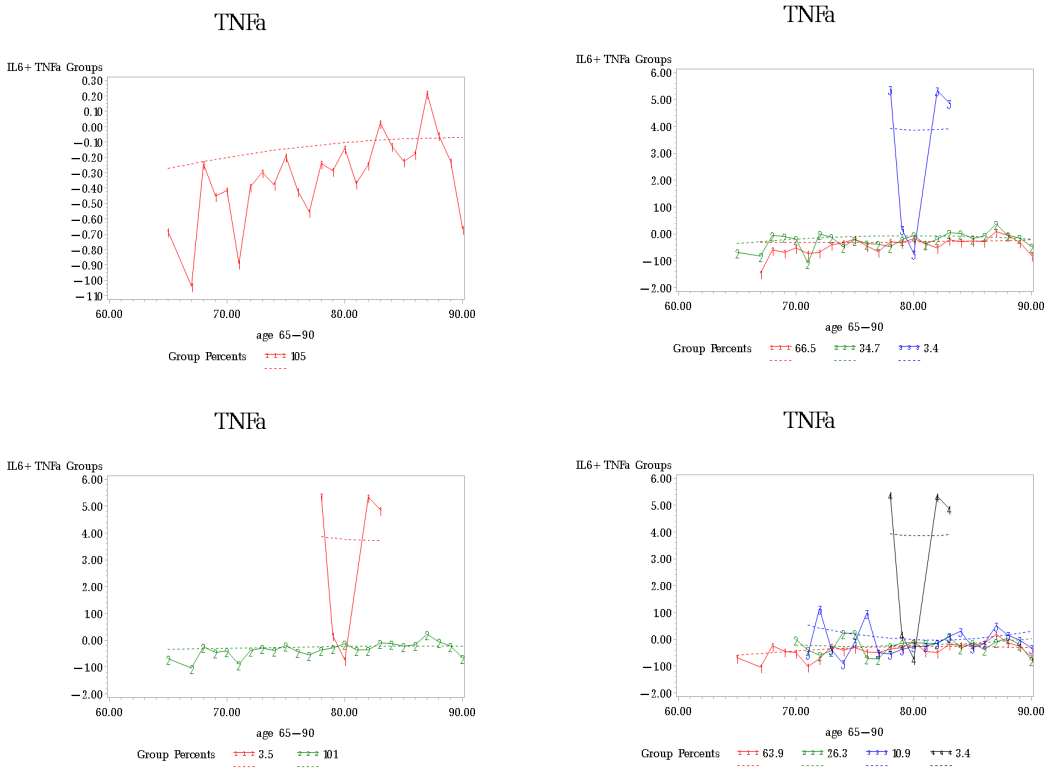


NOTE: PlGF trajectory plots based on single-trajectory GBTMs with 1, 2, 3, and 4 trajectory groups considered. PlGF: placental growth factor; GBTM: group-based trajectory model.

Figure 5. IL6 Group Number Selection Plots

NOTE: IL6 trajectory plots based on multi-trajectory GBTMs for the outcomes IL6 and TNFa with 1, 2, 3, and 4 trajectory groups considered. GBTM: group-based trajectory model; IL: interleukin, TNFa: tumor necrosis factor alpha.

Figure 6. TNFα Group Number Selection Plots



NOTE: TNFα trajectory plots based on multi-trajectory GBTMs for the outcomes IL6 and TNFα with 1, 2, 3, and 4 trajectory groups considered. GBTM: group-based trajectory model; IL: interleukin, TNFα: tumor necrosis factor alpha.

Table 2. Aβ42/40 and ptau181 Trajectory Order Reduction

			Iteration											
			1	2	3	4	5	6	7	8	9	10	11	12
Outcome	**Group**	**Parameter**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**
Aβ42/40	1	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Quadratic	0.0036	0.0036	0.0035	0.0037	0.0041	0.0038	0.0036	0.0038	0.0039	0.0052	0.0053	0.0052
	2	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.7087	0.7099	0.7107	0.6936	0.6712	0.6578	0.5331	0.5298	0.4638	0.525	0.5293	0.5364
		Quadratic	0.0106	0.0105	0.0104	0.0102	0.0091	0.0121	0.0135	0.0094	0.0095	0.0091	0.0097	0.012
	3	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.1783	0.1778	0.1813	0.1807	0.1795	0.1719	0.1303	0.1177	-	-	-	-
		Quadratic	0.1646	0.1643	0.1606	0.1605	0.1602	0.1852	0.1926	-	-	-	-	-
	4	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.2676	0.2671	0.2659	0.2717	0.2811	0.3045	-	-	-	-	-	-
		Quadratic	0.5158	0.5155	0.5071	0.5045	0.4879	-	-	-	-	-	-	-
ptau	1	Intercept	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0008	0.0008	0.0009
		Linear	0.0004	0.0004	0.0004	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0035	0.0035	0.0035
		Quadratic	0.1032	0.1033	0.1022	0.0973	0.0962	0.091	0.0866	0.0885	0.0869	-	-	-
	2	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.5624	0.563	0.5657	0.5997	-	-	-	-	-	-	-	-
		Quadratic	0.7009	0.6926	0.692	-	-	-	-	-	-	-	-	-
	3	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0814	0.0808	0.0654	0.0647	0.0601	0.0625	0.0576	0.0621	0.0655	0.0663	-	-
		Quadratic	0.5338	0.5353	-	-	-	-	-	-	-	-	-	-
	4	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.064	0.0624	0.0616	0.0595	0.0433	0.0467	0.0395	0.04	0.0406	0.0398	0.1099	-
		Quadratic	0.9555	-	-	-	-	-	-	-	-	-	-	-

NOTE: P-values for trajectory polynomial order reduction via manual backwards selection. Aβ: amyloid beta, ptau: phosphorylated tau.

Table 3. PlGF Trajectory Order Reduction

			Iteration					
			1	2	3	4	5	6
Outcome	**Group**	**Parameter**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**
PlGF	1	Intercept	0.7511	0.7517	0.4825	0.4798	0.4214	0.7598
		Linear	0.0415	0.0414	0.0422	0.044	0.0631	-
		Quadratic	0.7484	0.7476	-	-	-	-
	2	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0785	0.0771	0.0799	0.086	-	-
		Quadratic	0.7005	0.7158	0.717	-	-	-
	3	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Quadratic	0.8114	-	-	-	-	-
	4	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0295	0.0268	0.0268	0.0269	0.0334	0.0336
		Quadratic	0.0401	0.0395	0.0395	0.0396	0.0386	0.0387

NOTE: P-values for trajectory polynomial order reduction via manual backwards selection. PlGF: placental growth factor.

Table 4. IL6 and TNFa Trajectory Order Reduction

			Iteration										
			1	2	3	4	5	6	7	8	9	10	11
Outcome	**Group**	**Parameter**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**	**P-value**
IL6	1	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0598	0.0597	0.0725	0.1421	0.1462	0.1285	0.0689	0.0535	0.051	0.0493	0.0497
		Quadratic	0.0555	0.0555	0.0513	0.0405	0.0402	0.0408	0.0456	0.0466	0.0497	0.0489	0.0491
	2	Intercept	0.0188	0.0188	0.0242	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Quadratic	0.5847	0.5845	-	-	-	-	-	-	-	-	-
	3	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0004	0.0004	0.0004	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		Quadratic	0.0923	0.0923	0.0939	0.2252	0.2289	0.2588	0.2612	0.2646	-	-	-
	4	Intercept	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.1098	0.0595	0.0595	0.0619	0.0626	0.0622	0.0622	0.0622	0.063	0.0636	-
		Quadratic	0.8699	-	-	-	-	-	-	-	-	-	-
TNFa	1	Intercept	<0.0001	<0.0001	<0.0001	0.0017	0.0025	0.0065	0.0116	0.0099	0.0103	0.0079	0.0082
		Linear	0.5315	0.5313	0.5465	0.4714	0.4557	0.431	0.5063	-	-	-	-
		Quadratic	0.294	0.294	0.2682	0.2158	0.2134	0.324	-	-	-	-	-
	2	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.0127	0.0126	0.0122	0.0784	0.0852	0.1356	0.1377	0.0882	0.0909	-	-
		Quadratic	0.1032	0.1032	0.1028	0.2188	0.2342	-	-	-	-	-	-
	3	Intercept	0.6629	0.663	0.6105	0.7071	0.8796	0.879	0.8556	0.8356	0.858	0.9841	0.9803
		Linear	0.1246	0.1245	0.1321	0.6659	-	-	-	-	-	-	-
		Quadratic	0.2717	0.2716	0.2854	-	-	-	-	-	-	-	-
	4	Intercept	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		Linear	0.007	0.007	0.0071	0.0121	0.0136	0.0135	0.0135	0.0135	0.0136	0.0134	0.0145
		Quadratic	0.0035	0.0035	0.0035	0.0046	0.0048	0.0048	0.0048	0.0048	0.0049	0.0048	0.005

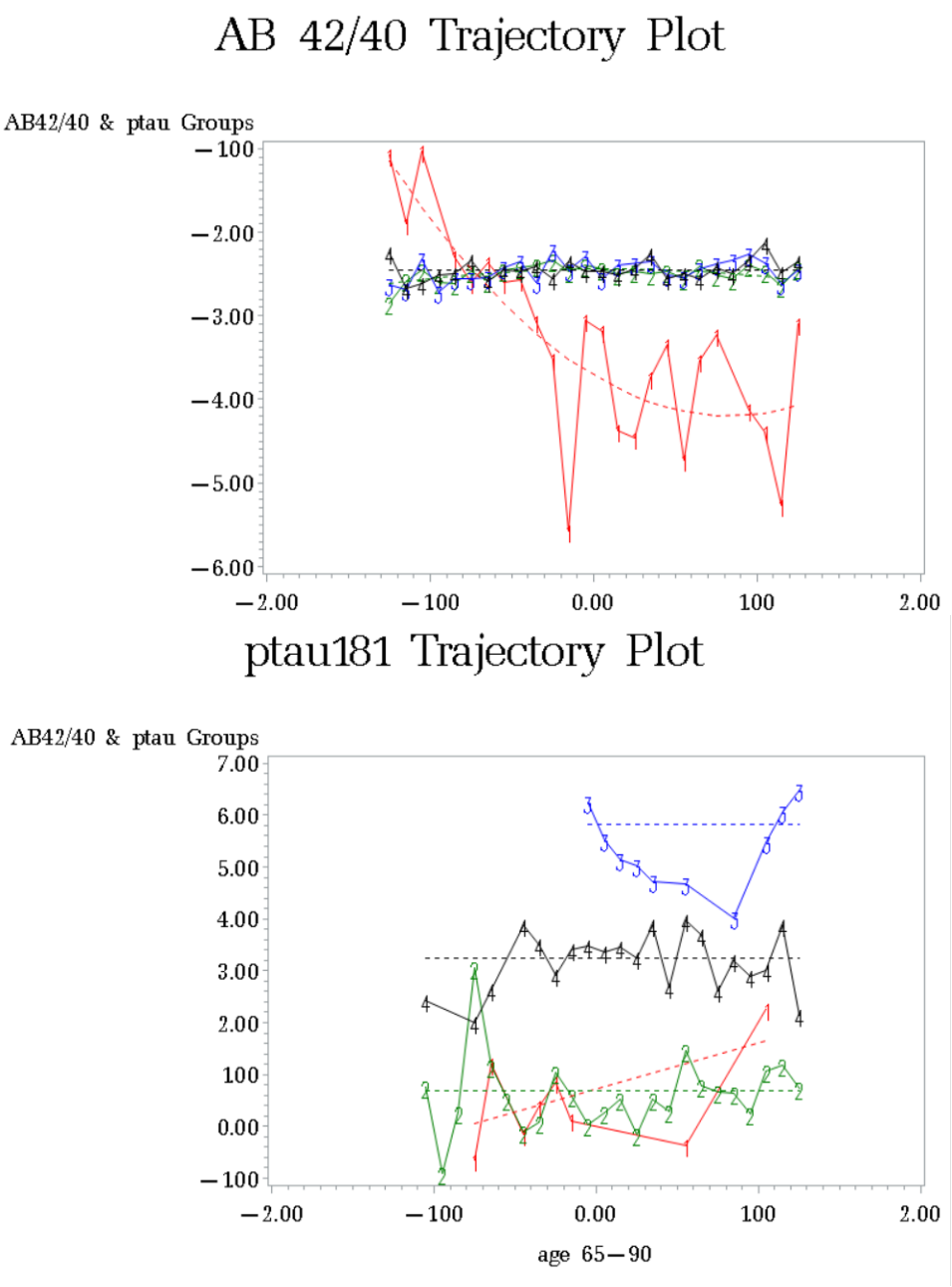
NOTE: P-values for trajectory polynomial order reduction via manual backwards selection. IL: interleukin; TNFa: tumor necrosis factor alpha

Table 5. P-values for Risk Factor Retention Tests

Outcome(s)	Risk Factor	Likelihood Ratio Tests	Wald Tests
Aβ42/40 + ptau	APOE	< .0001	0.3425
	Gender	0.8826	0.8801
	Education	< .0001	0.9716
	Race	0.0119	0.3101
PlGF	APOE	< .0001	0.6179
	Gender	< .0001	0.5365
	Education	< .0001	0.0370
	Race	0.2427	0.0368
IL6 + TNFa	APOE	< .0001	0.4435
	Gender	0.1703	0.8137
	Education	< .0001	0.0060
	Race	0.0027	0.0160

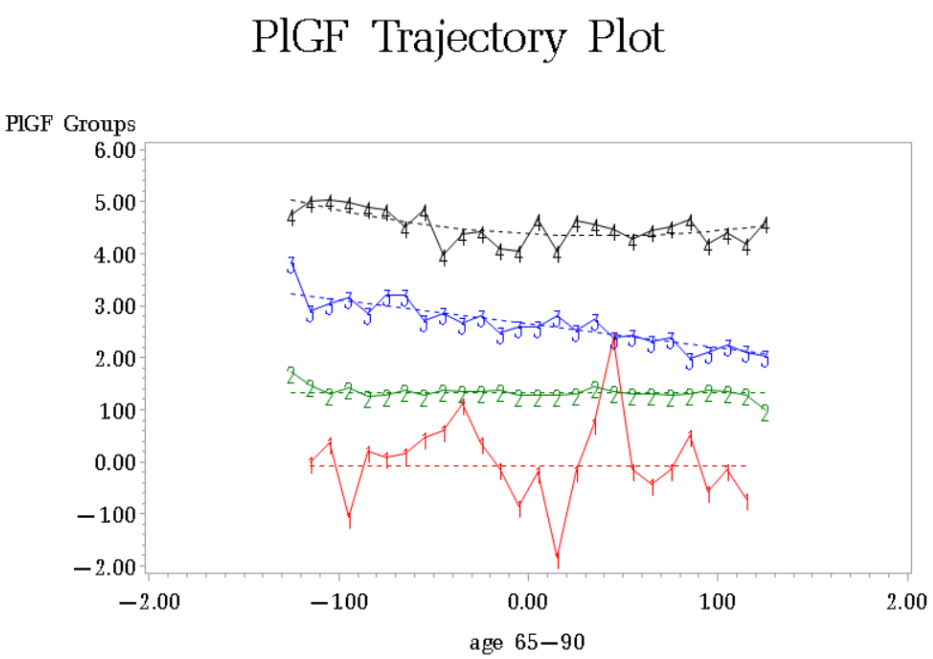
NOTE: Likelihood ratio test p-values result from Chi-Square test statistics (with 3 degrees of freedom) for twice the difference in the log likelihoods of the models with and without the specified risk factor. Wald test p-values result from Chi-Square test statistics (with 3 degrees of freedom) for a test of whether the risk factor coefficients for group membership are equal to 0. Aβ: amyloid beta; ptau: phosphorylated tau; PlGF: placental growth factor; IL: interleukin; TNFa: tumor necrosis factor alpha.

Figure 7. Aβ42/40 and ptau181 Final Trajectory Model Plots



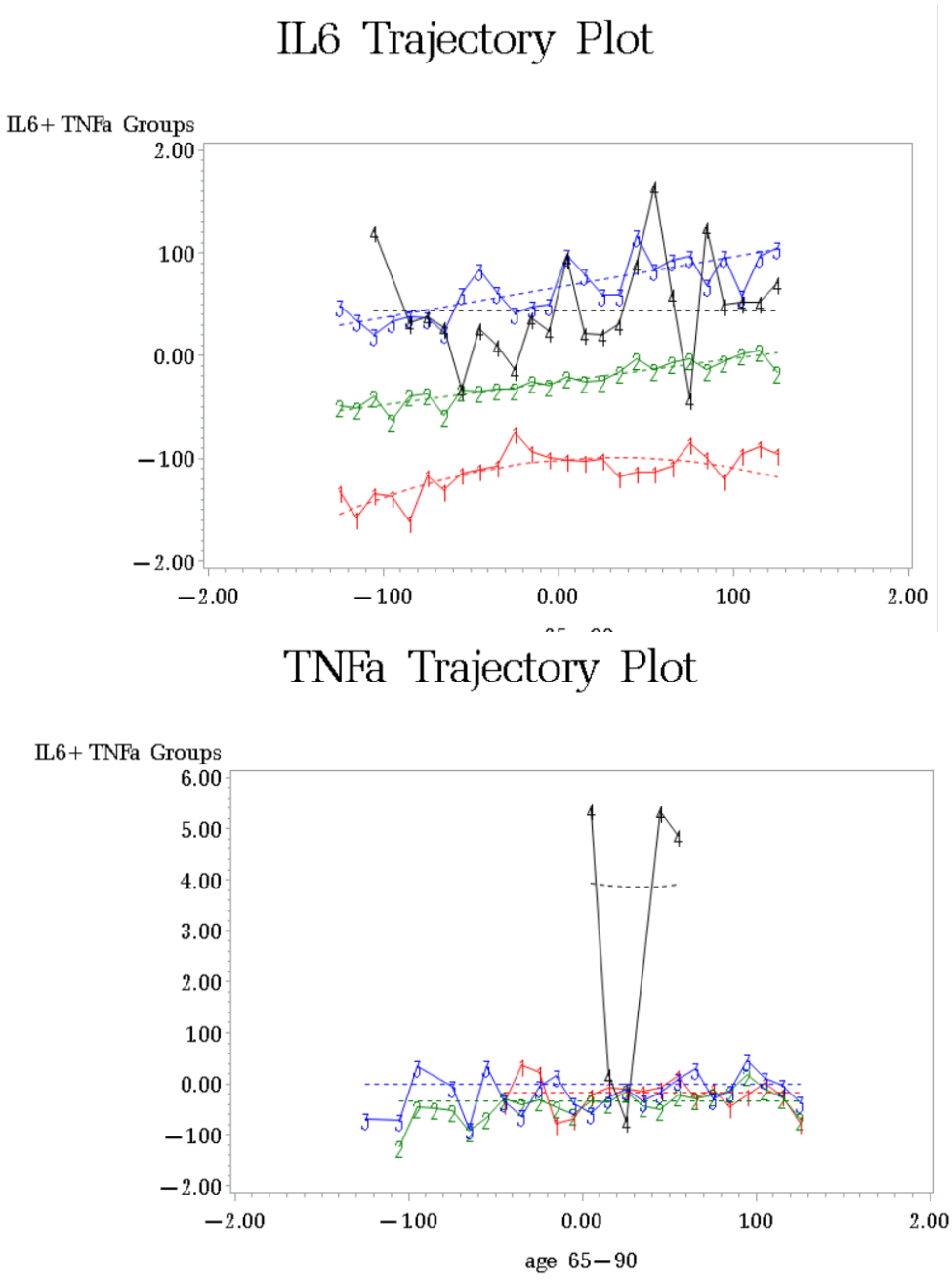
NOTE: Trajectory plots for Aβ42/40 and ptau181 from the final model (with reduced trajectory polynomial orders). Aβ: amyloid beta; ptau: phosphorylated tau.

Figure 8. PlGF Final Trajectory Model Plot



NOTE: Trajectory plot for the final PlGF model (with reduced trajectory polynomial orders). PlGF: placental growth factor.

Figure 9. IL6 and TNFα Final Trajectory Model Plots



NOTE: Trajectory plots for IL6 and TNFa from the final model (with reduced trajectory polynomial orders). IL: interleukin; TNFα: tumor necrosis factor alpha.

Table 6. Aβ42/40 and ptau181 Results Table

Sample Characteristics by Aβ42/40 & ptau181 Trajectory

Characteristics	Total Sample (N=916)	Trajectories 1 (N=16)	2 (N=663)	3 (N=84)	4 (N=153)	Test Statistic
Trajectory Characteristics						
Group Proportion		0.02	0.72	0.09	0.17	
Avg. Post. Prob.		0.794	0.811	0.913	0.861	
Baseline Characteristics						
Gender, N(%)						
Male	363(40)	7(44)	258(39)	38(45)	60(39)	$X^2_3 = 1.3708$
Female	553(60)	9(56)	405(61)	46(55)	93(61)	
Race, N(%)						
White	794(87)	14(88)	564(85)	75(89)	141(92)	$X^2_3 = 5.9711$
Other	122(13)	2(12)	99(15)	9(11)	12(8)	
APOE, N(%)						
Any e4 alleles	358(39)	8(50)	245(37)	37(44)	68(44)	$X^2_3 = 4.7809$
No e4 alleles	558(61)	8(50)	418(63)	47(56)	85(56)	
Education, mean(SD)	16.25(2.81)	15.12(2.53)	16.18(2.83)	16.68(2.87)	16.44(2.71)	$F_{3,912} = 1.864$
Cognitive Characteristics						
MMSE, mean(SD)						
First-Last	1.12(3.28)	0.93(1.67)	0.99(3.08)	0.65(2.94)	1.97(4.19)	$F_{3,845} = 4.126$
First-Last (1 vs 2, 3, 4)		0.93(1.67)		1.13(3.3)		$F_{1,847} = 0.051$
First-Last (1, 2 vs 3 vs 4)		0.99(3.06)		0.65(2.94)	1.97(4.19)	$F_{2,846} = 6.193$
Max-Last	1.65(3.15)	1.2(1.52)	1.49(2.96)	1.41(2.67)	2.51(4.1)	$F_{3,800} = 4.141$
Max-Last (1 vs 2, 3, 4)		1.2(1.52)		1.66(3.18)		$F_{1,802} = 0.31$
Max-Last (1, 2 vs 3 vs 4)		1.48(2.93)		1.41(2.67)	2.51(4.1)	$F_{2,801} = 6.154$
CDRGLOB, mean(SD)						
First-Last	-0.13(0.36)	-0.03(0.12)	-0.13(0.36)	-0.13(0.4)	-0.17(0.35)	$F_{3,879} = 1.077$
First-Last (1 vs 2, 3, 4)		-0.03(0.12)		-0.14(0.36)		$F_{1,881} = 1.334$
First-Last (1, 2 vs 3 vs 4)		-0.13(0.36)		-0.13(0.4)	-0.17(0.35)	$F_{2,880} = 1.051$
First-Max	-0.16(0.36)	-0.03(0.12)	-0.15(0.36)	-0.17(0.38)	-0.21(0.36)	$F_{3,879} = 2.054$
First-Max (1 vs 2, 3, 4)		-0.03(0.12)		-0.16(0.36)		$F_{1,881} = 2.024$
First-Max (1, 2 vs 3 vs 4)		-0.14(0.36)		-0.17(0.38)	-0.21(0.36)	$F_{2,880} = 2.272$
Clinical Diagnosis, N(%)						
Baseline Normal						
Normal	500(57)	4 (25)	351 (56)	56 (67)	89 (58)	$X^2_{15} = 38.409$
MCI	95(11)	0 (0)	72 (11)	5 (6)	18 (12)	
Demented	31(3)	1 (6)	20 (3)	1 (1)	9 (6)	
Baseline MCI						
MCI	103(12)	7 (44)	79 (13)	7 (8)	10 (7)	
Demented	33(4)	0 (0)	21 (3)	2 (2)	10 (7)	
Baseline Demented						
Demented	121(14)	4 (25)	87 (14)	13 (16)	17 (11)	

NOTE: Comparison of various characteristics across trajectories from the final GBTM for AB42/40 and ptau181. 33 individuals are missing cognitive diagnosis, all of which are in trajectory group 2, making 630 its effective N, and 833 the effective total sample size for clinical diagnosis. Any percentages that do not add to 100 are a result of rounding. Aβ: amyloid beta; ptau: phosphorylated tau; APOE*: apolipoprotein E; MMSE: Mini Mental State Exam; CDRGLOB: Clinical Dementia Rating global score; MCI: mild cognitive impairment; SD: standard deviation, N: sample size.*

Table 7. PlGF Results Table

Sample Characteristics by PlGF Trajectory

Characteristics	Total Sample (N=916)	Trajectories 1 (N=13)	2 (N=699)	3 (N=109)	4 (N=95)	Test Statistic
Trajectory Characteristics						
Group Proportion		0.01	0.76	0.12	0.10	
Avg. Post. Prob.		0.856	0.921	0.812	0.930	
Baseline Characteristics						
Gender, N(%)						
Male	363(40)	0(0)	268(38)	51(47)	44(46)	$X^2_3 = 13.13$
Female	553(60)	13(100)	431(62)	58(53)	51(54)	
Race, N(%)						
White	794(87)	6(46)	609(87)	96(88)	83(87)	$X^2_3 = 18.836$
Other	122(13)	7(54)	90(13)	13(12)	12(13)	
APOE, N(%)						
Any e4 alleles	358(39)	7(54)	266(38)	51(47)	34(36)	$X^2_3 = 4.6522$
No e4 alleles	558(61)	6(46)	433(62)	58(53)	61(64)	
Education, mean(SD)	16.25(2.81)	15.85(2.64)	16.03(2.81)	17.16(2.59)	16.87(2.82)	$F_{3,912} = 6.919$
Cognitive Characteristics						
MMSE, mean(SD)						
First-Last	1.12(3.28)	0.92(1.83)	1.17(3.38)	1.12(3.4)	0.82(2.52)	$F_{3,845} = 0.312$
Max-Last	1.65(3.15)	1.33(1.92)	1.71(3.32)	1.76(3.19)	1.1(1.56)	$F_{3,800} = 1.014$
CDRGLOB, mean(SD)						
First-Last	-0.13(0.36)	-0.19(0.56)	-0.12(0.34)	-0.23(0.48)	-0.11(0.31)	$F_{3,879} = 3.37$
First-Max	-0.16(0.36)	-0.19(0.56)	-0.15(0.33)	-0.24(0.48)	-0.14(0.33)	$F_{3,879} = 2.407$
Clinical Diagnosis, N(%)						
Baseline Normal						
Normal	500(57)	9(69)	373(56)	58(53)	60(63)	$X^2_{15} = 11.419$
MCI	95(11)	2(15)	71(11)	13(12)	9(9)	
Demented	31(3)	0	24(4)	5(5)	2(2)	
Baseline MCI						
MCI	103(12)	0	81(12)	13(12)	9(9)	
Demented	33(4)	0	21(3)	8(7)	4(4)	
Baseline Demented						
Demented	121(14)	2(15)	96(14)	12(11)	11(12)	

NOTE: Comparison of various characteristics across trajectories from the final GBTM for PlGF. 33 individuals are missing cognitive diagnosis, all of which are in trajectory group 2, making 666 its effective N, and 833 the effective total sample size for clinical diagnosis. Any percentages that do not add to 100 are a result of rounding. PlGF: placental growth factor; APOE*: apolipoprotein E; MMSE: Mini Mental State Exam; CDRGLOB: Clinical Dementia Rating global score; MCI: mild cognitive impairment; SD: standard deviation, N: sample size.*

Table 8. IL6 and TNFα Results Table

Sample Characteristics by IL6+TNFα Trajectory

Characteristics	Total Sample (N=916)	Trajectories 1 (N=145)	2 (N=632)	3 (N=111)	4 (N=28)	Test Statist
Trajectory Characteristics						
Group Proportion		0.16	0.69	0.12	0.03	
Avg. Post. Prob.		0.813	0.812	0.815	0.999	
Baseline Characteristics						
Gender, N(%)						
Male	363(40)	55(38)	261(41)	37(33)	10(36)	$X^2_3 = 2.928$
Female	553(60)	90(62)	371(59)	74(67)	18(64)	
Race, N(%)						
White	794(87)	109(75)	563(89)	97(87)	25(89)	$X^2_3 = 20.00$
Other	122(13)	36(25)	69(11)	14(13)	3(11)	
APOE, N(%)						
Any e4 alleles	358(39)	68(47)	232(37)	46(41)	12(43)	$X^2_3 = 5.641$
No e4 alleles	558(61)	77(53)	400(63)	65(59)	16(57)	
Education, mean(SD)	16.25(2.81)	16.5(2.7)	16.39(2.77)	15.23(3.01)	15.89(2.78)	$F_{3,912} = 6.08$
Cognitive Characteristics						
MMSE, mean(SD)						
First-Last	1.12(3.28)	0.61(3.17)	1.13(3.18)	1.75(4)	1.11(2.51)	$F_{3,845} = 2.40$
First-Last (1 vs 2 vs 3, 4)		0.61(3.17)	1.13(3.18)	1.61(3.74)		$F_{2,846} = 3.18$
First-Last (1, 2, 3 vs 4)			1.12(3.31)		1.11(2.51)	$F_{1,847} = 0.00$
Max-Last	1.65(3.15)	1.58(3.03)	1.55(3.03)	2.3(3.99)	1.54(2.36)	$F_{3,800} = 1.63$
Max-Last (1 vs 2 vs 3, 4)		1.58(3.03)	1.55(3.03)	2.13(3.71)		$F_{2,801} = 1.80$
Max-Last (1, 2, 3 vs 4)			1.65(3.18)		1.54(2.36)	$F_{1,802} = 0.03$
CDRGLOB, mean(SD)						
First-Last	-0.13(0.36)	-0.11(0.3)	-0.13(0.35)	-0.18(0.5)	-0.12(0.35)	$F_{3,879} = 0.76$
First-Last (1 vs 2 vs 3, 4)		-0.11(0.3)	-0.13(0.35)	-0.17(0.47)		$F_{2,880} = 0.93$
First-Last (1, 2, 3 vs 4)			-0.14(0.36)		-0.12(0.35)	$F_{1,881} = 0.02$
First-Max	-0.16(0.36)	-0.13(0.29)	-0.16(0.35)	-0.22(0.49)	-0.12(0.35)	$F_{3,879} = 1.33$
First-Max (1 vs 2 vs 3, 4)		-0.13(0.29)	-0.16(0.35)	-0.2(0.46)		$F_{2,880} = 1.27$
First-Max (1, 2, 3 vs 4)			-0.16(0.36)		-0.12(0.35)	$F_{1,881} = 0.24$
Clinical Diagnosis, N(%)						
Baseline Normal						
Normal	500(57)	81(56)	341(57)	61(55)	17(61)	$X^2_{15} = 16.73$
MCI	95(11)	19(13)	64(11)	11(10)	1(4)	
Demented	31(3)	8(6)	16(3)	5(5)	2(7)	
Baseline MCI						
MCI	103(12)	20(14)	67(11)	11(10)	5(18)	
Demented	33(4)	4(3)	27(5)	2(2)	0(0)	
Baseline Demented						
Demented	121(14)	13(9)	84(14)	21(19)	3(11)	

NOTE: Comparison of various characteristics across trajectories from the final GBTM for IL6 and TNFa. 33 individuals are missing cognitive diagnosis, all of which are in trajectory group 2, making 599 its effective N, and 833 the effective total sample size for clinical diagnosis. Any percentages that do not add to 100 are a result of rounding. IL: interleukin; TNFα: tumor necrosis factor; APOE*: apolipoprotein E; MMSE: Mini Mental State Exam; CDRGLOB: Clinical Dementia Rating global score; MCI: mild cognitive impairment; SD: standard deviation, N: sample size.*

Table 9. Post Hoc Cognitive Diagnosis Tests for Aβ42/40 and ptau181

Post Hoc Tests of Aβ42/40 & ptau181 Trajectories

Characteristics	Total Sample (N=916)	Trajectory 1 (N=16)	2 (N=663)	3 (N=84)	4 (N=153)	Test Statistic	P-Value
Post Hoc Cognitive Diagnosis Tests							
Baseline Clinical Diagnosis							
Normal	626(71)	5(31)	443(70)	62(74)	116(76)	$X^2_6 = 16.815$	0.00999
MCI	136(15)	7(44)	100(16)	9(11)	20(13)		
Demented	121(14)	4(25)	87(14)	13(15)	17(11)		
Final Clinical Diagnosis							
Normal	500(57)	4(25)	351(56)	56(67)	89(58)	$X^2_6 = 13.596$	0.03449
MCI	198(22)	7(44)	151(24)	12(14)	28(18)		
Demented	185(21)	5(31)	128(20)	16(19)	36(24)		
Progression (if possible)							
Yes	159(21)	1(8)	113(21)	8(11)	37(27)	$X^2_3 = 8.414$	0.03819
No	603(79)	11(92)	430(79)	63(89)	99(73)		
(Total)	(762)	(12)	(543)	(71)	(136)		

NOTE: Post hoc analysis of clinical cognitive diagnoses for Aβ42/40 and ptau181 trajectories following a significant test of clinical diagnosis transition. Aβ: amyloid beta; ptau: phosphorylated tau; MCI: mild cognitive impairment; N: sample size.

Table 10. Post Hoc Pairwise Comparisons

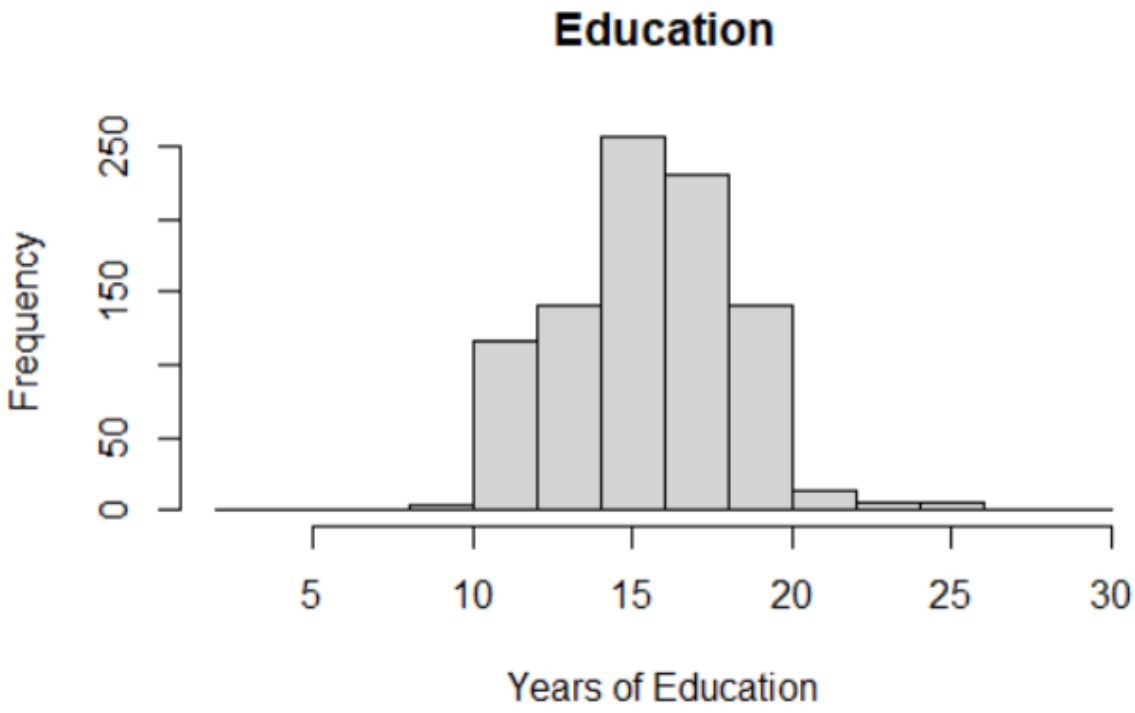
Post Hoc Pairwise Chi-Square Tests of Independence

	Comparison					
	Baseline Diagnosis		**Final Diagnosis**		**Progression**	
Trajectory Pairings	X^2_2	P-Value	X^2_2	P-Value	X^2_1	P-Value
Trajectory 1						
With Trajectory 2	12.142*	0.0023	6.081*	0.047	1.120*	0.2899
With Trajectory 3	13.518*	0.0012	10.983*	0.0041	0.091*	0.7629
With Trajectory 4	14.800*	0.0006	7.782*	0.0204	2.058*	0.1514
Trajectory 2						
With Trajectory 3	1.567	0.4568	4.663	0.0971	3.614	0.0573
With Trajectory 4	1.827	0.4011	2.468	0.2911	2.585	0.1079
Trajectory 3						
With Trajectory 4	1.092	0.5793	1.654	0.4373	6.965	0.0083

*NOTE: Post hoc pairwise comparisons of cognitive diagnosis characteristics (first identified in Table 9) for trajectories from the final Aβ42/40 and ptau181 model. X^2: Chi-square where subscript identifies degrees of freedom for test. *Tests with asterisk are not stable because >20% of cells have an expected size of <5* [178].

Supplemental Materials

Figure 1. Education Distribution



NOTE: Distribution of education for 916 individuals used in the analysis. Shape is not Normal, but it is unimodal and roughly symmetrical.

Chapter 5

Introduction

Throughout the course of this dissertation, we studied relatively new blood-based plasma biomarkers in relation to Alzheimer's Disease (AD) and measures of cognitive decline. These biomarkers have shown promise in similar literature [155-159], but do not have standard, established relationships with cognition and related outcomes. Chapter 1 presents a literature review of two statistical methodologies, nonparametric regression and mixture modeling, the latter of which lays the groundwork for the methodology used in the final analytic chapter. Chapter 2 is motivated by the need for a thorough examination of the relationships the biomarkers hold with other demographics and health conditions before the inclusion of cognitive factors. Chapter 3 attempts to examine a large number of biomarkers in relation to three cognitive outcomes, as well as clinical cognitive diagnosis, while accounting for established prognostic factors and identified influential health conditions. Finally, Chapter 4 utilizes mixture modeling through the extension of Group-Based Trajectory Modeling (GBTM) [160-161] to analyze whether the paths of five different biomarker measurements over a number of years can be divided into groups that significantly differ in a variety of cognitive assessments.

Summary

Literature Review

A brief review of nonparametric regression highlighted its utility as a method to handle real-world data that does not follow a known parametric distribution or is plagued with outliers. Both structure and estimates are derived from the data with nonparametric statistics, which requires generally larger sample sizes. A more careful analysis is required for nonparametric regression than its parametric counterpart, but its utility is not taken advantage of as often as it should be. Ultimately, nonparametric regression was not used for any analyses in this dissertation.

Our brief review of mixture modeling techniques laid the groundwork for the analysis in our penultimate chapter. At its most basic, mixture modeling is used to identify latent subgroups within a population. While it is theoretically plausible for there to be infinitely many mixture components, many real-world applications are limited to mixture models with a finite number of components. We covered a variety of methods for parameter estimation, including method of moments [36], maximum likelihood estimation [37-38], and expectation maximization algorithms [189]. While the modified likelihood ratio test (MLRT) improved upon the standard likelihood ratio test (LRT) for the estimation of mixtures [53], the D-test, which is distinct from the LRT and MLRT in its utilization of the L^2 distance, was shown to outperform the MLRT in some simulations [57]. Computational feasibility used to be a major burden but advancements in computational power and abilities have facilitated the use of mixture modeling methods for estimating mixtures beyond a mixture of 2 Normal densities to many more, possibly non-Normal, densities.

Chapter 2

We conducted a retrospective study on a cognitively Normal subset of the University of Kentucky Alzheimer's Disease Research Center (UKADRC) longitudinal cohort [70], using paired observations for individuals as the main and sensitivity cross-sectional analyses. Eleven different biomarkers and two ratios were evaluated. Each biomarker (and biomarker ratio) was individually analyzed via linear regression models, with covariates identified through sufficient adjustment sets of variables included in directed acyclic graphs (DAGs) representing the causal pathway from biomarkers to cognition [73]. In our sample, biomarkers were not largely associated with demographic or health factors, although some expected associations were observed (e.g., *APOE* with Aβ42/40 and tau/Aβ42). Sensitivity analyses found very similar results. These results encourage the use of plasma biomarkers in future dementia research without considerable worry of confounding due to demographic or medical conditions.

One major limitation of this analysis was the data quality. The instrument for measuring biomarkers was upgraded from the Quanterix SiMOA HD1 to the HDX between the running of our first and second paired years of data. Batch coefficients, which were nearly always significant, produced effect sizes that showed the majority of biomarker measurements from the HDX were predicted to be more than 1/3 of a standard deviation different than those from the HD1. While a difference in machine performance was not something we intended to study, our findings highlight the importance of thorough evaluation of data collection and measurement methods, especially for longitudinal data. Quanterix no longer advertises the SiMOA HD1 analyzer [72], but many of them may still be in use, and our study emphasizes the importance of clearly identifying machines used for data collection and the impact a machine update can have on longitudinal datasets.

Chapter 3

In this study we evaluated the usefulness of the same battery of plasma biomarkers, with the addition of phosphorylated tau (ptau181), in predicting current and future cognitive performance. Cognition was measured viz Mini Mental State Exam (MMSE) [17], California Verbal Learning Test long delay score (CVLT-LD) [119], and Trail Making Test B-A (Trails B-A) [127], and clinical cognitive diagnosis was used as an additional outcome in a post hoc analysis [70].

With individuals still contributing paired visits, four study designs were used to cover cross-sectional and longitudinal use of the data. Datasets 1 and 2 used all information from an individual's first and second visits, respectively, providing two cross-sectional analyses, the second of which can be thought of as a sensitivity analysis for the first. Dataset 3 used biomarker and covariate data from visit 1, and cognitive outcome data from visit 2, to evaluate use of the biomarkers in predicting future cognitive outcomes. Dataset 4 used covariate and cognitive outcome data from visit 2, while biomarkers were evaluated via change in values from visit 1 to visit 2, which may provide more insight into disease progression if change in biomarkers is predictive.

All variables identified in DAGs from the previous chapter, along with 5 prognostic factors, were run through a Lasso procedure for each cognitive outcome to reduce the number of variables and retain only those that influence the outcome. Residuals from linear models (with cognitive outcomes regressed on Lasso and prognostic variables) were used as the input for a partial least squares (PLS) analysis [129]. The PLS analysis contained only biomarker measurements and produced linear combinations of biomarkers that theoretically explain variation in the outcomes. Final models were run with the prognostic variables, Lasso identified variables, and first components (linear combinations of biomarkers) from the PLS procedure.

While it was our hope to identify predictive linear combinations of biomarkers, we were unable to identify any statistically significant combinations. While our statistical methodology is theoretically defensible, it was perhaps too convoluted to represent the relationships between the biomarkers and cognition. Our sample size was reasonable, but perhaps a larger group of individuals could facilitate the discovery of more clinically significant findings.

This study proved to be a lesson in negative findings. While positive findings are much more interesting, it is vitally important that researchers share negative findings. While our methodology did not prove to be useful here, PLS analysis is a valuable tool, and it may be very effective in slightly different situations. Though we were not able to add positive findings to the field of research, it is our hope that future research can be informed by, and even improve upon, our negative findings.

Chapter 4

Our final analytic study yielded the most interesting results of the dissertation. Using GBTMs [166], we examined age-related trajectories of five plasma biomarkers that have shown promise in dementia research. Our goal was to identify distinct, latent trajectories for each biomarker over time and examine whether they differ based on cognitive performance, which would be expected if they truly measure disease progression. Aβ42/40 and ptau were modeled via a multi-trajectory

GBTM, as were IL6 and TNFα. PlGF was given its own single-trajectory model. MMSE [116], Clinical Dementia Rating Global Score (CDRGLOB) [24], and clinical cognitive diagnosis [70] were evaluated across identified trajectory groups.

While PlGF and the IL6 and TNFα trajectories differed mainly on baseline risk factors (and only one cognitive measure), the trajectories for Aβ42/40 and ptau181 did not differ on any baseline risk factors but did differ on many cognitive characteristics. Aβ42/40 and ptau trajectories differed on many measured differences in MMSE over time, as well as clinical cognitive diagnosis transitions from first to last visit per individual. Further post hoc analysis of cognitive diagnosis across trajectories revealed that the trajectory groups significantly differed on baseline diagnosis, final diagnosis, and whether individuals who could progress to a worse diagnosis did. Pairwise comparisons of trajectory groups for these factors showed that most differences were likely driven by one trajectory group that was quite different from the rest. However, this trajectory group was very small and all of its pairwise Chi-Square tests presented warnings based on small cell sizes. While it is possible that this trajectory group is just an artifact of the data, it could also represent a unique set of individuals in the target population. More testing needs to be done to examine this relationship.

In all, this study adds valuable information to the field of AD plasma biomarker research. A sophisticated modeling technique was leveraged and found significant differences within the sample. The results suggest that Aβ42/40 and ptau181 are most related to MMSE and clinical diagnosis over time, and that PlGF and IL6 may be related to CDRGLOB and MMSE measures over time, respectively. While much future research needs to be done, this work provides a promising basis for the establishment of these relationships.

Strengths and Limitations

There were many positive aspects of our studies. The design of our data facilitated its use by both cross-sectional and longitudinal analysis, the latter of which we were able to leverage in

multiple ways (prediction of future outcomes, differences in paired observations, and GBTM longitudinal analysis). The individuals in our various samples are well characterized and seen repeatedly at the same facilities. A wide range of information is available for each observation (from the National Alzheimer's Coordinating Center (NACC) Uniform Dataset Version 3 (UDS3), as well as additional measures), and standard guidelines and protocol were followed for data collection and clinical cognitive diagnosis.

As with any study, there are drawbacks. One major limitation we faced was the quality of our data in Chapter 2. The sample consisted of paired observations for individuals from two sets of years and the first paired years of observations were not run on the same machine as the second paired year (and any years beyond used in future analyses). The old machine may be less reliable than the new one, and significant batch effects gave validity to our concern. While the analysis done in Chapter 2 contains some unreliable data, we are fortunate in that the cohort is ongoing and more data became available be the time Chapters 3 and 4 were performed, so the loss of the first paired year of data has minimal impact on subsequent analyses.

Finally, the external validity of the results from all three studies is quite limited. A defining feature of the UKADRC longitudinal cohort is that all individuals have consented to be followed through autopsy, restricting the data to individuals who may not be representative of the target population of elderly individuals. Additionally, the individuals in this cohort are very highly educated. Our external validity is perhaps the most restricted given the geographical considerations of the cohort. Because each individual included in the cohort has consented to autopsy, they must be within 4 hours of the University of Kentucky (in central Kentucky). This highly specific region cannot be said to represent the entire state and may not be generalizable to others.

Future Directions

Through both the examination of similar literature and the results we were able to find, it is clear that more validation is needed to establish any consistent relationships these biomarkers have

with AD and cognitive decline. It would be our desire in the future to leverage a larger, and particularly more diverse sample of individuals, perhaps from multiple states, to better assess these relationships in a sample that has more external validity. Releasing some of the restrictions our samples hold may facilitate the discovery of relationships that we could not detect.

More investigation should be done into linear combinations of biomarkers that may be influential in AD research, either though PLS or another method. While we were unable to find any useful biomarker combinations in Chapter 3, we still believe that these biomarkers do not work independently in the body and their use together will likely lean in the direction of their true relationships with cognition. Given the chance, we would repeat the analysis in Chapter 3 using a different dimension reduction technique or remove the Lasso and linear regression residuals steps and include all variables in the PLS analysis for comparison with our original results.

There are many desirable directions for future analysis informed by the findings in our final analytic chapter. Given the significant findings for Aβ42/40 and ptau trajectories, it could be very beneficial to examine the biomarkers with separate or dual-trajectory GBTMs to see what relationships remain when trajectory groups are not tied to the other biomarker. It would also be very interesting to evaluate Aβ40 and Aβ42 separately, possibly with a dual-trajectory model, to compare results of the individual biomarkers with the results from their ratio. Additionally, given that this method showed promise with our small number of studied biomarkers, we would want to see this analysis performed on an even wider range of biomarkers over time.

While it would be premature to state that blood-based plasma biomarkers should be collected and evaluated at regular intervals for elderly individuals based on our findings alone, coupled with results from similar studies it is possible that collection and evaluation of biomarkers could be quite useful in the future. If researchers continue to study these biomarkers, our results may contribute to the overall literature that leads to the establishment of firm relationships between these biomarkers and cognitive decline. Such firm relationships could facilitate an easily accessible form

of detection and/or prevention of cognitive decline for a disease that is known for its difficulty in both of those respects.

REFERENCES

1. Hoskin, Tanya. "Parametric and nonparametric: Demystifying the terms." Mayo Clinic. 2012.
2. Smith, Brian L., Billy M. Williams, and R. Keith Oswald. "Comparison of parametric and nonparametric models for traffic flow forecasting." Transportation Research Part C: Emerging Technologies 10.4 (2002): 303
3. Harwell, Michael R. "Choosing Between Parametric and Nonparametric Tests." Journal of Counseling and Development 67.1 (1988): 35-38. Web.
4. Hardle, Wolfgang. Applied Nonparametric Regression. Cambridge University Press, 1990.
5. Douglas M. Hawkins, and Jeffrey S. Simonoff. "Algorithm AS 282: High Breakdown Regression and Multivariate Estimation." Journal of the Royal Statistical Society. Series C (Applied Statistics) 42.2 (1993): 423-32. Web.
6. Ramsay, J. O., and B. W. Silverman. Applied Functional Data Analysis : Methods and Case Studies. New York: Springer, 2002. Print. Springer Ser. in Statistics.
7. Richard CHARNIGO, and Cidambi SRINIVASAN. "Self-consistent Estimation of Mean Response Functions and Their Derivatives." Canadian Journal of Statistics 39.2 (2011): 280-99. Web.
8. Mroz, Thomas A., and Timothy H. Savage. "Overfitting and biases in nonparametric kernel regressions using cross-validated bandwidths: a cautionary note." Unpublished manuscript, Department of Economics, University of North Carolina, Chapel Hill (1999).
9. Geenens, Gery. "Curse of Dimensionality and Related Issues in Nonparametric Functional Regression." Statistics Surveys 5 (2011): 30-43. Web.
10. Martins-Filho, Carlos, and Yao, Feng. "Kernel-based Estimation of Semiparametric Regression in Triangular Systems." Economics Letters 115.1 (2012): 24-27. Web.
11. Wu, Yi, Wu, Yi, Wang, Xuejun, and Wang, Xuejun. "A Note on the Consistency for the Estimators of Semiparametric Regression Model." Statistical Papers (Berlin, Germany) 59.3 (2018): 1117-130. Web.
12. Clark, R. M. "Non-Parametric Estimation of a Smooth Regression Function." Journal of the Royal Statistical Society. Series B, Methodological 39.1 (1977): 107-13. Web.
13. Cleveland, William S. "Robust Locally Weighted Regression and Smoothing Scatterplots." Journal of the American Statistical Association 74.368 (1979): 829-36. Web.
14. Hastie, Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition (2nd ed. 2009.). Springer New York. https://doi.org/10.1007/978-0-387-84858-7
15. Loader, Clive. Local Regression and Likelihood. New York: Springer, 1999. Print. Statistics and Computing.
16. Stromberg, Arnold J, Hössjer, Ola, and Hawkins, Douglas M. "The Least Trimmed Differences Regression Estimator and Alternatives." Journal of the American Statistical Association 95.451 (2000): 853-64. Web.

17. Rousseeuw, Peter J. "Least Median of Squares Regression." Journal of the American Statistical Association 79.388 (1984): 871-80. Web.
18. Watson, Geoffrey S. "Smooth Regression Analysis." Sankhya. Series A 26.4 (1964): 359-72. Web.
19. Nadaraya, E. A. "On Estimating Regression." Theory of Probability and Its Applications 9.1 (1964): 141-42. Web.
20. Altman, N. S. "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." The American Statistician 46.3 (1992): 175-85. Web.
21. Delaigle, Aurore. "Nonparametric Kernel Methods with Errors-in-Variables: Constructing Estimators, Computing Them, and Avoiding Common Mistakes." Australian & New Zealand Journal of Statistics 56.2 (2014): 105-24. Web.
22. Mesquita, D.P.P, Gomes, J.P.P, and Souza Junior, A.H. "Epanechnikov Kernel for Incomplete Data." Electronics Letters 53.21 (2017): 1408-410. Web.
23. Lizhong Ding, and Shizhong Liao. "An Approximate Approach to Automatic Kernel Selection." IEEE Transactions on Cybernetics 47.3 (2017): 554-65. Web.
24. Belkin, Mikhail, Hsu, Daniel, Ma, Siyuan, and Mandal, Soumik. "Reconciling Modern Machine-learning Practice and the Classical Bias-variance Trade-off." Proceedings of the National Academy of Sciences - PNAS 116.32 (2019): 15849-5854. Web.
25. Hazelton, Martin L. "Variable Kernel Density Estimation." Australian & New Zealand Journal of Statistics 45.3 (2003): 271-84. Web.
26. Bischoff, W, Heck, B, Howind, J, and Teusch, A. "A Procedure for Testing the Assumption of Homoscedasticity in Least Squares Residuals: A Case Study of GPS Carrier-phase Observations." Journal of Geodesy 78.7 (2005): 397-404. Web.
27. Hall, Peter, Park, Byeong U, and Samworth, Richard J. "Choice of Neighbor Order in Nearest-neighbor Classification." The Annals of Statistics 36.5 (2008): 2135-152. Web.
28. Smit, S.K, and Eiben, A.E. "Comparing Parameter Tuning Methods for Evolutionary Algorithms." 2009 IEEE Congress on Evolutionary Computation (2009): 399-406. Web.
29. Homrighausen, Darren, and McDonald, Daniel J. "A Study on Tuning Parameter Selection for the High-dimensional Lasso." (2016). Web.
30. Hurvich, Clifford M, Simonoff, Jeffrey S, and Tsai, Chih-Ling. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion." Journal of the Royal Statistical Society. Series B, Statistical Methodology 60.2 (1998): 271-93. Web.
31. Aydın, Dursun & Tuzemen, Seref. (2012). Smoothing Parameter Selection Problem in Nonparametric Regression Based on Smoothing Spline: A Simulation Study. Journal of Applied Sciences. 12. 636-644. 10.3923/jas.2012.636.644.
32. Charnigo R, Hall B, Srinivasan C. A Generalized Cp Criterion for Derivative Estimation. Technometrics. 2011;53(3):238-253. doi: https://doi.org/10.1198/tech.2011.09147
33. Stone, Charles J. "Optimal Rates of Convergence for Nonparametric Estimators." The Annals of Statistics 8.6 (1980): 1348-360. Web.

34. Stone, Charles J. "Consistent Nonparametric Regression." The Annals of Statistics 5.4 (1977): 595-620. Web.
35. Redner RA, Walker HF. Mixture Densities, Maximum Likelihood and the EM Algorithm. SIAM Review. 1984;26(2):195-239. Accessed December 22, 2021. https://www.jstor.org/stable/2030064
36. Pearson K. Contributions to the Mathematical Theory of Evolution. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1894;185(0):71-110. doi:https://doi.org/10.1098/rsta.1894.0003
37. Godambe VP. An Optimum Property of Regular Maximum Likelihood Estimation. The Annals of Mathematical Statistics. 1960;31(4):1208-1211. doi: https://doi.org/10.1214/aoms/1177705693
38. Aitchison J, Silvey SD. Maximum-Likelihood Estimation Procedures and Associated Tests of Significance. Journal of the Royal Statistical Society: Series B (Methodological). 1960;22(1):154-171. doi:https://doi.org/10.1111/j.2517-6161.1960.tb00362.x
39. Hosmer DW. On MLE of the parameters of a mixture of two normal distributions when the sample size is small. Communications in Statistics. 1973;1(3):217-227. doi:https://doi.org/10.1080/03610927308827019
40. Yakowitz S. A Consistent Estimator for the Identification of Finite Mixtures. The Annals of Mathematical Statistics. 1969;40(5):1728-1735. Accessed June 25, 2023. https://www.jstor.org/stable/2239559
41. Yakowitz S. Unsupervised learning and the identification of finite mixtures. IEEE Transactions on Information Theory. 1970;16(3):330-338. doi: https://doi.org/10.1109/tit.1970.1054442
42. Odell PL, Chhikara R. Estimation of a large area crop acreage inventory using remote sensing technology in Annual Report: Statistical Theory and Methodology for Remote Sensing Data Analysis, Rep. NASA/JSC-09703, Univ. of Texas at Dallas, Dallas, TX, 1975
43. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological). 1977;39(1):1-38. https://www.jstor.org/stable/2984875
44. Hasselblad V. Estimation of Parameters for a Mixture of Normal Distributions. Technometrics. 1966;8(3):431-444. doi: https://doi.org/10.1080/00401706.1966.10490375
45. Behboodian J. On a mixture of normal distributions. Biometrika. 1970;57(1):215-217. doi:https://doi.org/10.1093/biomet/57.1.215
46. Wolfe JH. PATTERN CLUSTERING BY MULTIVARIATE MIXTURE ANALYSIS. Multivariate Behavioral Research. 1970;5(3):329-350. doi: https://doi.org/10.1207/s15327906mbr0503_6
47. Day N. Estimating the Components of a Mixture of Normal Distributions. Biometrika. 1969;56(3):463-463. doi: https://doi.org/10.2307/2334652
48. Casella G, Berger RL. Statistical Inference. Brooks/Cole Cengage Learning; 2017.
49. K. Ghosh J., Kumar S Pranab. On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. Institute of

Statistics mimeo series. 1984;(1467). Accessed June 25, 2023. http://www.lib.ncsu.edu/resolver/1840.4/3493
50. Chen H, Chen J. The likelihood ratio test for homogeneity in finite mixture models. Canadian Journal of Statistics. 2001;29(2):201-215. doi: https://doi.org/10.2307/3316073
51. Beran R. Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements. Journal of the American Statistical Association. 1988;83(403):687-687. doi: https://doi.org/10.2307/2289292
52. McLachlan GJ. On Bootstrapping the Likelihood Ratio Test Stastistic for the Number of Components in a Normal Mixture. Applied Statistics. 1987;36(3):318. doi:https://doi.org/10.2307/2347790
53. Chen H, Chen J, Kalbfleisch JD. A modified likelihood ratio test for homogeneity in finite mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63(1):19-29. doi:https://doi.org/10.1111/1467-9868.00273
54. Neyman J. Optimal asymptotic tests of composite hypotheses. Probability and Statistics. Published online January 1, 1959:213-234.
55. Kocherlakota S, Kocherlakota K. Neyman's C(α) test and Rao's efficient score test for composite hypotheses. Statistics & Probability Letters. 1991;11(6):491-493. doi: https://doi.org/10.1016/0167-7152(91)90113-6
56. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika. 1987;74(1):33-43. doi:https://doi.org/10.1093/biomet/74.1.33
57. Charnigo R, Sun J. Testing Homogeneity in a Mixture Distribution via the L2 Distance Between Competing Models. Journal of the American Statistical Association. 2004;99(466):488-498. doi:https://doi.org/10.1198/016214504000000494
58. "Population Distribution in the United States in 2020, by Generation." Statista, Statista Research Department, June 2021, www.statista.com/statistics/296974/us-population-share-by-generation/.
59. Bureau, U.S. Census. "2020 Census Will Help Policymakers Prepare for the Incoming Wave of Aging Boomers." Census.gov, 8 Oct. 2021, www.census.gov/library/stories/2019/12/by-2030-all-baby-boomers-will-be-age-65-or-older.html.
60. "Facts and Figures." Alzheimer's Association, 1 Mar. 2021, www.alz.org/alzheimers-dementia/facts-figures.
61. Nelson, Peter T., et al. "Alzheimer's Disease Is Not 'brain Aging': Neuropathological, Genetic, and Epidemiological Human Studies." Acta Neuropathologica, vol. 121, no. 5, Springer-Verlag, 2011, pp. 571–87, https://doi.org/10.1007/s00401-011-0826-y.
62. "Preventing Alzheimer's Disease: What Do We Know?" National Institute on Aging, U.S. Department of Health and Human Services, 24 Sept. 2018, https://www.nia.nih.gov/health/preventing-alzheimers-disease-what-do-we-know.
63. Karanth, Shama, et al. "Prevalence and Clinical Phenotype of Quadruple Misfolded Proteins in Older Adults." JAMA Neurology, vol. 77, no. 10,

American Medical Association, 2020, pp. 1299–307, https://doi.org/10.1001/jamaneurol.2020.1741.
64. Shaw, Leslie M., et al. “Cerebrospinal Fluid Biomarker Signature in Alzheimer’s Disease Neuroimaging Initiative Subjects.” Annals of Neurology, vol. 65, no. 4, Wiley Subscription Services, Inc., A Wiley Company, 2009, pp. 403–13, https://doi.org/10.1002/ana.21610.
65. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263-269. doi:10.1016/j.jalz.2011.03.005
66. Hampel, Harald et al. “State-of-the-art of lumbar puncture and its place in the journey of patients with Alzheimer's disease.” Alzheimer's & dementia : the journal of the Alzheimer's Association (2021): n. pag.
67. Karikari TK, Ashton NJ, Brinkmalm G, et al. Blood phospho-tau in Alzheimer disease: analysis, interpretation, and clinical utility. Nat Rev Neurol. 2022;18(7):400-418. doi:10.1038/s41582-022-00665-2
68. Li, Danni, and Michelle M. Mielke. “An Update on Blood-Based Markers of Alzheimer’s Disease Using the SiMoA Platform.” Neurology and Therapy, vol. 8, no. Suppl 2, Springer Healthcare, 2019, pp. 73–82, https://doi.org/10.1007/s40120-019-00164-5.
69. Mielke MM, Dage JL, Frank RD, et al. Performance of plasma phosphorylated tau 181 and 217 in the community. Nature Medicine. 2022;28(7):1398-1405. doi:https://doi.org/10.1038/s41591-022-01822-2
70. Schmitt FA, Nelson PT, Abner E, et al. University of Kentucky Sanders-Brown healthy brain aging volunteers: donor characteristics, procedures and neuropathology. Curr Alzheimer Res. 2012;9(6):724-733. doi:10.2174/156720512801322591
71. Quanterix. “HD-X: Brochure.” Ultra-Sensitive Biomarker Detection, 2019, go.quanterix.com/HDXBrochure.
72. “Immunoassay Platforms, Kits & Services.” Quanterix, 4 Jan. 2022, https://www.quanterix.com/products-and-services/.
73. Shrier, Ian, and Robert W. Platt. “Reducing Bias through Directed Acyclic Graphs.” BMC Medical Research Methodology, vol. 8, no. 1, BMC, 2008, pp. 70–70, https://doi.org/10.1186/1471-2288-8-70
74. Roses, Allen D., and Ann M. Saunders. “APOE Is a Major Susceptibility Gene for Alzheimer’s Disease.” Current Opinion in Biotechnology, vol. 5, no. 6, Elsevier Ltd, 1994, pp. 663–67, https://doi.org/10.1016/0958-1669(94)90091-4.
75. Abner, Erin L., et al. “Mild Cognitive Impairment: Statistical Models of Transition Using Longitudinal Clinical Data.” International Journal of Alzheimer’s Disease, 2012, pp. 291920–29, https://doi.org/10.1155/2012/291920.
76. Feng, Changyong, et al. “Log-Transformation and Its Implications for Data Analysis.” Shanghai Jingshen Yixue, vol. 26, no. 2, 2014, pp. 105–09, https://doi.org/10.3969/j.issn.1002-0829.2014.02.
77. Feng, Changyong, et al. “Log Transformation: Application and Interpretation in Biomedical Research.” Statistics in Medicine, vol. 32, no. 2, Blackwell Publishing Ltd, 2013, pp. 230–39, https://doi.org/10.1002/sim.5486.

78. Peña, Edsel A., and Elizabeth H. Slate. “Global Validation of Linear Model Assumptions.” Journal of the American Statistical Association, vol. 101, no. 473, Taylor & Francis, 2006, pp. 341–54, https://doi.org/10.1198/016214505000000637.
79. Emerson, Scott S. “Use of Ratios and Logarithms in Statistical Regression Models.” Emerson Statistics, 22 Jan. 2014, http://www.emersonstatistics.com/GeneralMaterials/Ratios%20and%20Logarithms.pdf.
80. Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
81. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
82. Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.
83. Hadley Wickham and Evan Miller (2019). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.2.0. https://CRAN.R-project.org/package=haven
84. Hadley Wickham and Jennifer Bryan (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl
85. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
86. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4(43):1686. doi:10.21105/joss.01686
87. Simrén J, Leuzy A, Karikari TK, et al. The diagnostic and prognostic capabilities of plasma biomarkers in Alzheimer's disease. Alzheimers Dement. 2021;17(7):1145-1156. doi:10.1002/alz.12283
88. Brickman AM, Manly JJ, Honig LS, et al. Plasma p-tau181, p-tau217, and other blood-based Alzheimer’s disease biomarkers in a multiethnic, community study. Alzheimers Dement. 2021;17(8):1353-1364. doi:10.1002/alz.12301
89. De Meyer, Steffi, et al. “Comparison of ELISA- and SIMOA-Based Quantification of Plasma Aβ Ratios for Early Detection of Cerebral Amyloidosis.” Alzheimer’s Research & Therapy, vol. 12, no. 1, BioMed Central, 2020, pp. 162–162, https://doi.org/10.1186/s13195-020-00728-w.
90. Deniz K, Ho CCG, Malphrus KG, et al. Plasma biomarkers of Alzheimer’s disease in African Americans. J Alzheimers Dis. 2021;79(1):323-334. doi:10.3233/JAD-200828
91. Lasseter HC, Provost AC, Chaby LE, Daskalakis NP, Haas M, Jeromin A. Cross-platform comparison of highly sensitive immunoassay technologies for cytokine markers: platform performance in post-traumatic stress disorder and Parkinson’s disease. Cytokine X. 2020;2(2):100027. doi:10.1016/j.cytox.2020.100027
92. Meyer PF, Ashton NJ, Karikari TK, et al. Plasma p-tau231, p-tau181, PET biomarkers, and cognitive change in older adults. Ann Neurol. 2022;91(4):548-560. doi:10.1002/ana.26308

93. Wolf, Frank, et al. “Plasma Tau, Neurofilament Light Chain and Amyloid-Beta Levels and Risk of Dementia; a Population-Based Cohort Study.” Brain (London, England : 1878), vol. 143, no. 4, Oxford University Press, 2020, pp. 1220–32, https://doi.org/10.1093/brain/awaa054.
94. Lambe, Susan, et al. “Perceptions, Knowledge, Incentives, and Barriers of Brain Donation Among African American Elders Enrolled in an Alzheimer’s Research Program.” The Gerontologist, vol. 51, no. 1, OXFORD UNIV PRESS INC, 2011, pp. 28–38, https://doi.org/10.1093/geront/gnq063.
95. Dementia. Mayo Clinic. Published October 12, 2022. Accessed November 15, 2022. http://www.mayoclinic.org/diseases-conditions/ dementia/symptoms-causes/syc-20352013
96. Woloshin S, Kesselheim AS. What to know about the Alzheimer drug aducanumab (Aduhelm). JAMA Internal Medicine. 2022;182(8):892. doi:10.1001/jamainternmed.2022.1039
97. Aducanumab approved for treatment of Alzheimer’s disease. Alzheimer’s Disease and Dementia. Accessed June 5, 2023. https://www.alz.org/alzheimers-dementia/treatments/aducanumab.
98. Lecanemab approved for treatment of early Alzheimer’s disease. Alzheimer’s Disease and Dementia. Accessed June 5, 2023. https://www.alz.org/alzheimers-dementia/treatments/lecanemab-leqembi.
99. Canady VA. FDA approves new treatment for Alzheimer’s disease. Mental Health Weekly. 2023;33(3):6-7. doi:10.1002/mhw.33505
100. 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021;17(3):327-406. doi:10.1002/alz.12328
101. Zvěřová M. Alzheimer's disease and blood-based biomarkers - potential contexts of use. Neuropsychiatr Dis Treat. 2018;14:1877-1882. Published 2018 Jul 20. doi:10.2147/NDT.S172285
102. Hampel H, Shaw LM, Aisen P, et al. State-of-the-art of lumbar puncture and its place in the journey of patients with Alzheimer’s disease. Alzheimers Dement. 2022;18(1):159-177. doi:10.1002/alz.12372
103. Li D, Mielke MM. An update on blood-based markers of Alzheimer’s disease using the SiMoA platform. Neurol Ther. 2019;8(2):73-82. doi:10.1007/s40120-019-00164-5
104. Schneider JA. Neuropathology of Dementia Disorders. Continuum (Minneap Minn). 2022;28(3):834-851. doi:10.1212/CON.0000000000001137
105. Chatterjee P, Pedrini S, Ashton NJ, et al. Diagnostic and prognostic plasma biomarkers for preclinical Alzheimer's disease. Alzheimers Dement. 2022;18(6):1141-1154. doi:10.1002/alz.12447
106. de Wolf F, Ghanbari M, Licher S, et al. Plasma tau, neurofilament light chain and amyloid-β levels and risk of dementia; a populationbased cohort study. Brain. 2020;143(4):1220-1232. doi:10.1093/ brain/awaa054
107. Weekman EM, Wilcock DM. Matrix Metalloproteinase in Blood-Brain Barrier Breakdown in Dementia. J Alzheimers Dis. 2016;49(4):893-903. doi:10.3233/JAD-150759
108. Karikari TK, Pascoal TA, Ashton NJ, et al. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction

modelling study using data from four prospective cohorts. Lancet Neurol. 2020;19(5):422-433. doi:10.1016/S1474-4422(20)30071-5
109. Zetterberg H. Blood-based biomarkers for Alzheimer's disease-An update. J Neurosci Methods. 2019;319:2-6. doi:10.1016/j.jneumeth.2018.10.025
110. De Meyer S, Schaeverbeke JM, Verberk IMW, et al. Comparison of ELISA- and SIMOA-based quantification of plasma Aβ ratios for early detection of cerebral amyloidosis. Alzheimers Res Ther. 2020;12(1):162. doi:10.1186/s13195-020-00728-w
111. Tanase DM, Gosav EM, Radu S, et al. Arterial Hypertension and Interleukins: Potential Therapeutic Target or Future Diagnostic Marker?. Int J Hypertens. 2019;2019:3159283. Published 2019 May 2. doi:10.1155/2019/3159283
112. Leuzy A, Mattsson-Carlgren N, Palmqvist S, Janelidze S, Dage JL, Hansson O. Blood-based biomarkers for Alzheimer's disease. EMBO Mol Med. 2022;14(1):e14408. doi:10.15252/emmm.202114408
113. Chen YR, Liang CS, Chu H, et al. Diagnostic accuracy of blood biomarkers for Alzheimer's disease and amnestic mild cognitive impairment: A meta-analysis. Ageing Res Rev. 2021;71:101446. doi:10.1016/j.arr.2021.101446
114. Syrjanen JA, Campbell MR, Algeciras-Schimnich A, et al. Associations of amyloid and neurodegeneration plasma biomarkers with comorbidities. Alzheimers Dement. 2022;18(6):1128-1140. doi:10.1002/alz. 12466
115. Estepp TG, Charnigo RJ, Abner EL, et al. Associations of potential ADRD plasma biomarkers in cognitively normal volunteers. Alzheimer’s Dement. 2023;1-9. https://doi.org/10.1002/alz.13000
116. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6
117. Besser L, Kukull W, Knopman DS, et al. Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. Alzheimer Dis Assoc Disord. 2018;32(4):351-358. doi:10.1097/WAD.0000000000000279
118. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc. 1992;40(9):922-935. doi:10.1111/j.1532-5415.1992.tb01992.x
119. Lundervold AJ, Wollschläger D, Wehling E. Age and sex related changes in episodic memory function in middle aged and older adults. Scand J Psychol. 2014;55(3):225-232. doi:10.1111/sjop.12114
120. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI)
121. Alexander MP, Stuss DT, Fansabedian N. California Verbal Learning Test: performance by patients with focal frontal and non-frontal lesions. Brain. 2003;126(Pt 6):1493-1503. doi:10.1093/brain/awg128
122. Delis DC, Kramer JH, Kaplan E, Ober BA. California Verbal Learning Test--Second Edition. PsycTESTS Dataset. Published online 1987. doi:https://doi.org/10.1037/t15072-000
123. Dobbs BM, Shergill SS. How effective is the Trail Making Test (Parts A and B) in identifying cognitively impaired drivers?. Age Ageing. 2013;42(5):577-581. doi:10.1093/ageing/aft073

124. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol. 2004;19(2):203-214. doi:10.1016/S0887-6177(03)00039-8
125. Terada S, Sato S, Nagao S, et al. Trail making test B and brain perfusion imaging in mild cognitive impairment and mild Alzheimer's disease. Psychiatry Res. 2013;213(3):249-255. doi:10.1016/j.pscychresns.2013.03.006
126. Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. Arch Clin Neuropsychol. 2008;23(2):129-137. doi:10.1016/j.acn.2007.11.005
127. Arbuthnott K, Frank J. Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. J Clin Exp Neuropsychol. 2000;22(4):518-528. doi:10.1076/1380-3395(200008)22:4;1-0;FT518
128. Stark J, Palombo DJ, Hayes JP, et al. Partial Least Squares Analysis of Alzheimer's Disease Biomarkers, Modifiable Health Variables, and Cognition in Older Adults with Mild Cognitive Impairment. J Int Neuropsychol Soc. 2022;28(8):781-789. doi:10.1017/S1355617721001041
129. Hastie T, Friedman J, Tisbshirani R. Chapter 3: Linear Methods for Regression. In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer; 2017.
130. Friedman J, Tibshirani R, Hastie T (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent.” Journal of Statistical Software, 33(1), 1–22. doi:10.18637/jss.v033.i01.
131. Hastie T, Friedman J, Tisbshirani R. Chapter 4: Linear Methods for Classification. In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer; 2017.
132. Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.
133. Obianom O. Introduction to r2symbols, An R Package for adding symbols to Rmarkdown and Shiny Apps. R-project.org. Published October 23, 2022. Accessed November 29, 2022. https://cran.r-project.org/web/packages/r2symbols/vignettes/introduction_r_symbols.html
134. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer International Publishing; 2016
135. Auguie B. Miscellaneous Functions for “Grid” Graphics [R package gridExtra version 2.3]. Published online 2017. Accessed November 16, 2022. https://CRAN.R-project.org/package=gridExtra
136. Read Excel Files [R package readxl version 1.4.1]. Published online 2022. Accessed November 16, 2022. https://CRAN.R-project.org/ package=readxl
137. Import and Export “SPSS”, “Stata” and “SAS” Files [R package haven version 2.5.1]. Published online 2022. Accessed November 16, 2022. https://CRAN.R-project.org/package=haven
138. Segovia F, Górriz JM, Ramírez J, Salas-González D, Álvarez I. Early diagnosis of Alzheimer’s disease based on partial least squares and support vector machine.

Expert Systems with Applications. 2013;40(2):677-683. doi:10.1016/j.eswa.2012.07.071
139. Wang Y, Ibrahim JG, Zhu H. Partial least squares for functional joint models with applications to the Alzheimer's disease neuroimaging initiative study. Biometrics. 2020;76(4):1109-1119. doi:10.1111/biom.13219
140. Chiesa PA, Houot M, Vergallo A, et al. Association of brain network dynamics with plasma biomarkers in subjective memory complainers. Neurobiol Aging. 2020;88:83-90. doi:10.1016/j.neurobiolaging.2019.12.017
141. Lemercier P, Vergallo A, Lista S, et al. Association of plasma Aβ40/Aβ42 ratio and brain Aβ accumulation: testing a whole-brain PLS-VIP approach in individuals at risk of Alzheimer's disease. Neurobiol Aging. 2021;107:57-69. doi:10.1016/j.neurobiolaging.2021.07.005
142. Wang DC, Sun CH, Liu LY, et al. Serum fatty acid profiles using GC-MS and multivariate statistical analysis: potential biomarkers of Alzheimer's disease. Neurobiol Aging. 2012;33(6):1057-1066. doi:10.1016/j.neurobiolaging.2010.09.013
143. Doorduijn AS, Visser M, van de Rest O, et al. Associations of AD Biomarkers and Cognitive Performance with Nutritional Status: The NUDAD Project. Nutrients. 2019;11(5):1161. Published 2019 May 23. doi:10.3390/nu11051161
144. Nordlund A, Göthlin M, Wallin A. Vascular disease, Alzheimer's disease biomarkers and cognition in mild cognitive impairment: additive or synergetic effects?. Dement Geriatr Cogn Disord. 2011;32(4):250-256. doi:10.1159/000334653
145. Howlett J, Hill SM, Ritchie CW, Tom BDM. Disease Modelling of Cognitive Outcomes and Biomarkers in the European Prevention of Alzheimer's Dementia Longitudinal Cohort. Front Big Data. 2021;4:676168. Published 2021 Aug 20. doi:10.3389/fdata.2021.676168
146. Nair AS. Publication bias - Importance of studies with negative results!. Indian J Anaesth. 2019;63(6):505-507. doi:10.4103/ija.IJA_142_19
147. Joober R, Schmitz N, Annable L, Boksa P. Publication bias: what are the challenges and can they be overcome?. J Psychiatry Neurosci. 2012;37(3):149-152. doi:10.1503/jpn.120065
148. Kalaria RN, Ballard C. Overlap between pathology of Alzheimer disease and vascular dementia. Alzheimer Dis Assoc Disord. 1999;13 Suppl 3:S115-S123. doi:10.1097/00002093-199912003-00017
149. DeTure MA, Dickson DW. The Neuropathological Diagnosis of Alzheimer’s Disease. Molecular Neurodegeneration. 2019;14(1):1-18. doi:https://doi.org/10.1186/s13024-019-0333-5
150. Raz L, Knoefel J, Bhaskar K. The neuropathology and cerebrovascular mechanisms of dementia. J Cereb Blood Flow Metab. 2016;36(1):172-186. doi:10.1038/jcbfm.2015.164
151. de Souza LC, Sarazin M, Teixeira-Júnior AL, Caramelli P, Santos AE, Dubois B. Biological markers of Alzheimer's disease. Arq Neuropsiquiatr. 2014;72(3):227-231. doi:10.1590/0004-282x20130233
152. Jack CR Jr, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic

biomarkers. Lancet Neurol. 2013;12(2):207-216. doi:10.1016/S1474-4422(12)70291-0
153. Schneider JA. Vascular and Alzheimer's pathology: A mixed affair. Alzheimer's & Dementia. 2020;16(S4). doi:https://doi.org/10.1002/alz.038813
154. Fani L, Ahmad S, Ikram MK, Ghanbari M, Ikram MA. Immunity and amyloid beta, total tau and neurofilament light chain: Findings from a community-based cohort study. Alzheimers Dement. 2021;17(3):446-456. doi:10.1002/alz.12212
155. Li Y, Schindler SE, Bollinger JG, et al. Validation of Plasma Amyloid-β 42/40 for Detecting Alzheimer Disease Amyloid Plaques. Neurology. 2022;98(7):e688-e699. doi:10.1212/WNL.0000000000013211
156. Meng J, Lei P. Plasma pTau181 as a biomarker for Alzheimer's disease. MedComm (2020). 2020;1(1):74-76. Published 2020 May 22. doi:10.1002/mco2.1
157. Hinman JD, Elahi F, Chong D, et al. Placental growth factor as a sensitive biomarker for vascular cognitive impairment [published online ahead of print, 2023 Feb 23]. Alzheimers Dement. 2023;10.1002/alz.12974. doi:10.1002/alz.12974
158. Custodero C, Ciavarella A, Panza F, et al. Role of inflammatory markers in the diagnosis of vascular contributions to cognitive impairment and dementia: a systematic review and meta-analysis. Geroscience. 2022;44(3):1373-1392. doi:10.1007/s11357-022-00556-w
159. Decourt B, Lahiri DK, Sabbagh MN. Targeting Tumor Necrosis Factor Alpha for Alzheimer's Disease. Curr Alzheimer Res. 2017;14(4):412-425. doi:10.2174/1567205013666160930110551
160. Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. Psychological Methods. 1999;4(2):139-157. doi:https://doi.org/10.1037/1082-989x.4.2.139
161. Nagin DS. Group-Based Trajectory Modeling: An Overview. Annals of Nutrition and Metabolism. 2014;65(2-3):205-210. doi:https://doi.org/10.1159/000360229
162. Chapman KR, Bing-Canar H, Alosco ML, et al. Mini Mental State Examination and Logical Memory scores for entry into Alzheimer's disease trials. Alzheimer's Research & Therapy. 2016;8(1). doi:https://doi.org/10.1186/s13195-016-0176-z
163. Coding Guidebook for Initial Visit Packet.; 2013. Accessed June 16, 2023. https://files.alz.washington.edu/documentation/uds3-ivp-guidebook.pdf
164. Data Element Dictionary for Initial Visit Visit Packet.; 2015. Accessed June 16, 2023. https://files.alz.washington.edu/documentation/uds3-ivp-ded.pdf
165. O'Bryant SE, Waring SC, Cullum CM, et al. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. Arch Neurol. 2008;65(8):1091-1095. doi:10.1001/archneur.65.8.1091
166. Nagin D. Group-Based Modeling of Development. Harvard University Press; 2005.
167. Joens BL, Nagin DS, Roeder K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Methods & Research. 2001;29(3):374-393. doi:https://doi.org/10.1177/0049124101029003005

168. Koscik RL, Betthauser TJ, Jonaitis EM, et al. Amyloid duration is associated with preclinical cognitive decline and tau PET. Alzheimers Dement (Amst). 2020;12(1):e12007. Published 2020 Feb 13. doi:10.1002/dad2.12007
169. Jonaitis EM, Janelidze S, Cody KA, et al. Plasma phosphorylated tau 217 in preclinical Alzheimer's disease. Brain Commun. 2023;5(2):fcad057. Published 2023 Mar 6. doi:10.1093/braincomms/fcad057
170. Betthauser TJ, Koscik RL, Jonaitis EM, et al. Chronicity of amyloid: Methods for estimating amyloid onset and clinical research applications. Alzheimer's and Dementia. 2020;16(S5). doi:https://doi.org/10.1002/alz.039782
171. Ding X, Charnigo RJ, Schmitt FA, Kryscio RJ, Abner EL. Evaluating trajectories of episodic memory in normal cognition and mild cognitive impairment: Results from ADNI. Ginsberg SD, ed. PLOS ONE. 2019;14(2):e0212435. doi:https://doi.org/10.1371/journal.pone.0212435
172. McLachlan GJ, Lee SX, Rathnayake SI. Finite Mixture Models. Annual Review of Statistics and Its Application. 2019;6(1):355-378. doi:https://doi.org/10.1146/annurev-statistics-031017-100325
173. Jones BL, Nagin DS. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Methods & Research. 2007;35(4):542-571. doi:https://doi.org/10.1177/0049124106292364
174. Du, Gaixin, "CLUSTERING HOSPITAL PERFORMANCE USING GROUP-BASED MULTI-TRAJECTORY MODELING WITH SINGULAR BAYESIAN INFORMATION CRITERION" (2023). Theses and Dissertations--Epidemiology and Biostatistics. 38. https://uknowledge.uky.edu/epb_etds/38
175. Arrandale VH, Mieke Koehoorn, MacNab YC, Kennedy SM. How to use SAS® Proc Traj and SAS® Proc Glimmix in respiratory epidemiology. Published online December 1, 2006. doi:https://doi.org/10.14288/1.0048205
176. Coelho CA, Marques FJ, Jorge N, Nunes C. On the distribution of the likelihood ratio test of independence for random sample size — a computational approach. Journal of computational and applied mathematics. 2021;404(113394):113394-113394. doi:https://doi.org/10.1016/j.cam.2021.113394
177. Cuevas A, Febrero M, Fraiman R. An anova test for functional data. Computational Statistics & Data Analysis. 2004;47(1):111-122. doi:https://doi.org/10.1016/j.csda.2003.10.021
178. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod. 2017;42(2):152-155. doi:10.5395/rde.2017.42.2.152
179. Jones B. traj: group-based modeling of longitudinal data. www.andrew.cmu.edu. Published 2023. Accessed June 21, 2023. https://www.andrew.cmu.edu/user/bjones/download.htm
180. Wickham H. plyr: Tools for Splitting, Applying and Combining Data. R-Packages. Published November 11, 2022. https://cran.r-project.org/web/packages/plyr/index.html
181. Anisimova M, Bielawski JP, Yang Z. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Molecular Biology and Evolution. 2001;18(8):1585-1592. doi:https://doi.org/10.1093/oxfordjournals.molbev.a003945

182. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012;10(2):486-489. doi:10.5812/ijem.3505
183. Assumptions for ANOVA. Real-statistics.com. Published May 21, 2022. https://real-statistics.com/one-way-analysis-of-variance-anova/assumptions-anova/
184. Ali MM, Sharma SC. Robustness to nonnormality of regression F-tests. Journal of Econometrics. 1996;71(1-2):175-205. doi:https://doi.org/10.1016/0304-4076(94)01700-x
185. Odoi B, Twumasi-Ankrah S, Samita S, Al-Hassan S. The Efficiency of Bartlett's Test using Different forms of Residuals for Testing Homogeneity of Variance in Single and Factorial Experiments-A Simulation Study. Scientific African. 2022;17(e01323):e01323. doi:https://doi.org/10.1016/j.sciaf.2022.e01323
186. Stephanie. Likelihood-Ratio Tests (Probability and Mathematical Statistics). Statistics How To. Published September 15, 2016. https://www.statisticshowto.com/likelihood-ratio-tests/
187. Liu W, Wang Y, Xie W. Fisher information matrix, Rao test, and Wald test for complex-valued signals and their applications. Signal Processing. 2014;94:1-5. doi:https://doi.org/10.1016/j.sigpro.2013.06.032
188. Mésidor M, Rousseau MC, O'Loughlin J, Sylvestre MP. Does group-based trajectory modeling estimate spurious trajectories? BMC Medical Research Methodology. 2022;22(1). doi:https://doi.org/10.1186/s12874-022-01622-9
189. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological). 1977;39(1):1-38. https://www.jstor.org/stable/2984875
190. Drton M, Plummer M. A Bayesian information criterion for singular models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2017;79(2):323-380. doi:https://doi.org/10.1111/rssb.12187

VITA

Taylor Estepp

Education

- University of Kentucky
  Master of Science, Statistics, 2018
- Eastern Kentucky University
  Bachelor of Science, Statistics, 2016

Professional Positions

- Graduate Research Assistant
  Sanders-Brown Center on Aging, May 2019 – present
- Graduate Research Assistant
  UK Department of Health Management and Policy, June 2018 – May 2019
- Graduate Instructor
  UK Department of Statistics, June 2017 – May 2018
- Graduate Teaching Assistant
  UK Department of Statistics, August 2016 – June 2017

Scholastic Honors

- Student representative for UK College of Public Health Strategic Planning Committee. Summer/Fall 2019.
- Awarded 2nd Place for Ph.D. student poster presentations. Poster based on publication of Chapter 2 manuscript (cited in Acknowledgements). UK College of Public Health Public Health Showcase. April 2023.

Publications

- Estepp TG, Abner EL, Fleming ST. Chapter 12 - Randomized Clinical Trials, Managerial Epidemiology: Cases and Concepts, 4th Edition by Dr. Steven T. Fleming. September 29, 2020.
- Estepp TG, Charnigo RJ, Abner EL, et al. Associations of potential ADRD plasma biomarkers in cognitively normal volunteers. Alzheimer's Dement. 2023;1-9.